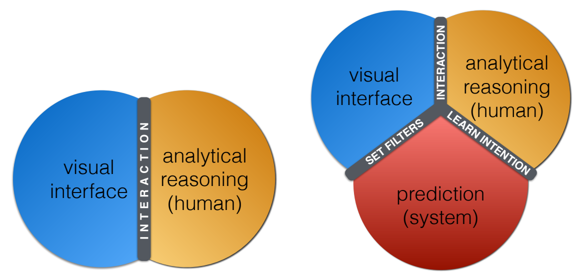
Accepted at the Hawaii International Conference on System Sciences (HICSS-49).

Mixed-Initiative for Big Data: The Intersection of Human + Visual Analytics + Prediction

Stephen Makonin*, Daniel McVeigh†, Wolfgang Stuerzlinger‡*, Khoa Tran*, Fred Popowich*†
*Computing Science, †Cognitive Science, ‡School of Interactive Arts + Technology – Simon Fraser University
Email: {smakonin, dmcveigh, w.s, khoa_tran, popowich}@sfu.ca

***Abstract*—Existing surveys in visual analytics focus on the importance of the topic. However, many do not discuss the increasingly critical area of mixed-initiative systems. In this survey we discuss the importance of research in mixed-initiative systems and how it is different from visual analytics and other research fields. We present the conceptual architecture of a mixed-initiative visual analytics system (MIVAS) and the five key components that make up MIVASs (data wrangling, alternative discovery and comparison, parametric interaction, history tracking and exploration, and system agency and adaptation), which forms our main contribution. We compare and contrast different research that claims to be mixed-initiative against MIVASs and show how there is still a considerable amount of work that needs to be accomplished before any system can truly be mixed-initiative.**

***Index Terms*—mixed-initiative, visual analytics, big data, prediction, human computer interaction**



(a) Visual Analytics (VA) (b) Mixed-Initiative (MI)

Fig. 1. In (a), VA is the science of analytical reasoning facilitated by interactive visual interfaces [1]. In (b), mixed initiative approaches either the computer or the human can take initiative and decide what to do next [wiktionary.org].

I. Introduction

Overall, people who are trained and tasked to reason analytically about data (experts) are good at it. However, when these analytical experts are faced with overwhelmingly large amounts of (big) data their capacity for cognitive processing *all* of the presented data in its raw form proves quite limited. As humans rely largely on vision for most interaction, it is often beneficial for their of analytical reasoning if this data is summarized and visualized in various different ways, such as charts, networks, glyphs, and icons. However, static (or closed) visualizations do not enable further exploration for possible insights and to generate new hypotheses in a way that is intuitive and matches the human analytic reasoning process well. Adding interactive elements, where a human can readily perform queries through a visual interface, or fine-tune the visualization in real-time, e.g., zoom, scroll, add/remove details/filters, and add/remove data, we enter the field of visual analytics (see Figure 1(a)).

Yet, visual analytics (VA) has generally not used machine learning techniques within visual interaction to assist and enhance human analytical reasoning [2]–[4]. Examples for potential assistance include having the system prompt the human when it detects potential issues, identifies uncertainties within datasets, or filters unwanted results. This kind of system is often referred to as mixed-initiative (MI) [5]–[9], where a human user and a predictive system, or machine-agent, work in tandem to initiate actions towards accomplishing a common goal (see Figure 1(b) for a visual definition). In VA, this may include setting output filters, modifying parameters for mining algorithms and mapping models, exploring alternative mining algorithms, identifying patterns for large classifications, and setting visualization parameters based on the agents understanding of the task at hand. The predictive agent can then communicate results to the user through a visualization or more explicitly via dialogue; reflexively, the agent may learn and adapt based on input from direct dialogue with the user and from an awareness and memory of the user interacting with the VA system (see Figure 1(b)).

In this survey we look in more detail at past research in the various areas that comprise MI (Section II). We then define the necessary functionality that a complete mixed-initiative VA system (MIVAS) must include (Section III) and critically compare and contrast existing MI systems to these criteria: data wrangling (Section IV), alternative discovery and comparison (Section V), parametric interaction (Section VI), history tracking and exploration (Section VII), and system agency and adaptation (Section VIII). We identified these five components through our in depth analysis of existing literature. With the exception of data wrangling, most of these concepts already exist but are not well identified or defined. This is an attempt to do so while encapsulating them in a framework that shows how each component is a feedback loop that reuses many existing systems (Figure 2) – but in different combinations/ways. Table I summarizes our critical review of the current state-of-the-art research by comparing each research project or commercial package with the necessary functionality of MI systems. We then discuss the broader implications of MI systems and their critical part in VA and big data research (Section IX). Lastly, conclusions (Section X).

II. Other Relevant Surveys

A. Visual Analytics Surveys

As the amount of data increases, the task to analyze and make sense of data and to pull insight from such larger and larger datasets become increasingly difficult [10]. VA takes advantage of two powerful human capacities to assist in the analytics process; the human eye, supporting the largest sensory bandwidth into the brain, and rapid reasoning skills, specifically rapid object recognition and classification [1]. Effectively representing datasets using highly salient visual features can enable an analyst to quickly review and explore large quantities of data, facilitating the discovery of relevant insight and hypotheses [11] – even measuring the confidence of these discoveries [12].

Many visualization techniques have been created to help the human analytical process [13]. Such techniques include: visualizing time-varying volumetric data [14], [15]; comparing one time series to another [16]; grouping data [17]; or ranking data variables over time or different attributes [18], [19]. Interestingly, much of this work does not focus on intelligent and adaptive interactions with the VA system.

Some surveys focus on a range of differentiable features in VA software [20], [21]. Although some mention the need for MI systems [3], [4], none actually survey the MI field. Other work has surveyed the core concepts of VA and human collaboration [22]. Further surveys within VA have focused on visualization and analytical functionalities of available software [20], [21], [23], [24]. A survey on intelligent agents in data analytics focuses primarily on hard-coded intelligence, and includes only a single example of a MI system [25]. A recent survey on tools and resources in modern analytics encourages the integration of original data and metadata to create software which can adapt to particular intentions or insights derived from the user interaction with the analytics system [26].

B. Predictive Analytics Surveys

Predictive analytics attempts to identify future events based on previous observations using different machine learning algorithms [27]. For the purposes of MI, predictive analytics can be used to predict future human intentions (i.e. the intended use of the system) as well as to predict which modelling parameters or attributes need to be filtered, set, or adjusted. Additionally, predictive analytics can learn from the task output of a human and propose an optimized version (e.g. the same 3D model but with a much lower polygonal count), or provide alternative analysis (e.g. try different learning algorithms to find one that yields better prediction). Recently [28]–[30] there is a focus on the algorithms used, where the algorithms performance is evaluated based on accuracy of prediction. Neither the visualization of the prediction nor the interaction techniques with said prediction are discussed.

Similar to VA where *interaction* is the bond between the visualization and the human, in MI two other forms of interactive (or communicative) bonding are required: interaction between data visualization and the prediction system, and interactive negotiation between the human and the prediction system.

C. Contribution

Our survey does draw on many concepts discussed in previous studies, such as [3], which discusses the possible benefits of MI software. Yet, there are currently very few examples within the field of VA.

Our focus is in surveying examples of MI research within the field of VA, identifying various means of implementing a MIVAS, and identifying under-utilized elements of adaptive MI agents for future research. The definition of MIVASs and the identification of its five key components forms our main contribution.

III. Defining Mixed-Initiative

A MI framework is defined as an efficient and natural collaboration between multiple agents directed towards completing a goal or negotiating a solution to a problem [31]. In software design, this interactive work effort may comprise the collaboration between a human and any number of machine-agents, or, in the case of complex-situation planning software, a collaboration between multiple machine-agents [32]. The advantage in splitting a workload is not merely to reduce the individual efforts involved in accomplishing a task, but to make the best use of individual strengths within a joint activity. For example, MI planning software can include collaboration between two machine-agents, each responsible for a specific component of the planning process [33].

Generally, MI collaborations occur between a human and a machine-agent, which plays to a more well defined division of strengths, namely the division between human and machine intelligence. Human intelligence provides expert domain knowledge and observational reasoning to a task, and also ultimately generates context-specific meaning to the results derived by computation, a concept referred to as *sense-making*. Machine intelligence is the product of two major factors, rapid access to large volumes of memory and bias-free information processing [34]. Bias-free reasoning can be powerful in data analysis. However, it would be a mistake to regards the bias-laden reasoning of a human, who (possibly) has a rich socio-cultural understanding of a given domain, merely as a human fault. Rather, both forms of reasoning should be applied in their appropriate problem solving contexts. Further, there is natural flexibility to human cognition, allowing us to quickly incorporate new information and adapt our reasoning accordingly. This capacity may be more effectively utilized in conjunction with a machine agent who can identify, possibly through pattern recognition, when a user has become *stuck* in a redundant or fruitless exploration, and suggest new alternatives [35].

In the next five sections we will discuss each part of functionality necessary for a complete mixed-initiative visual analytics system (MIVAS): data wrangling (Section IV), alternative discovery and comparison (Section V), parametric

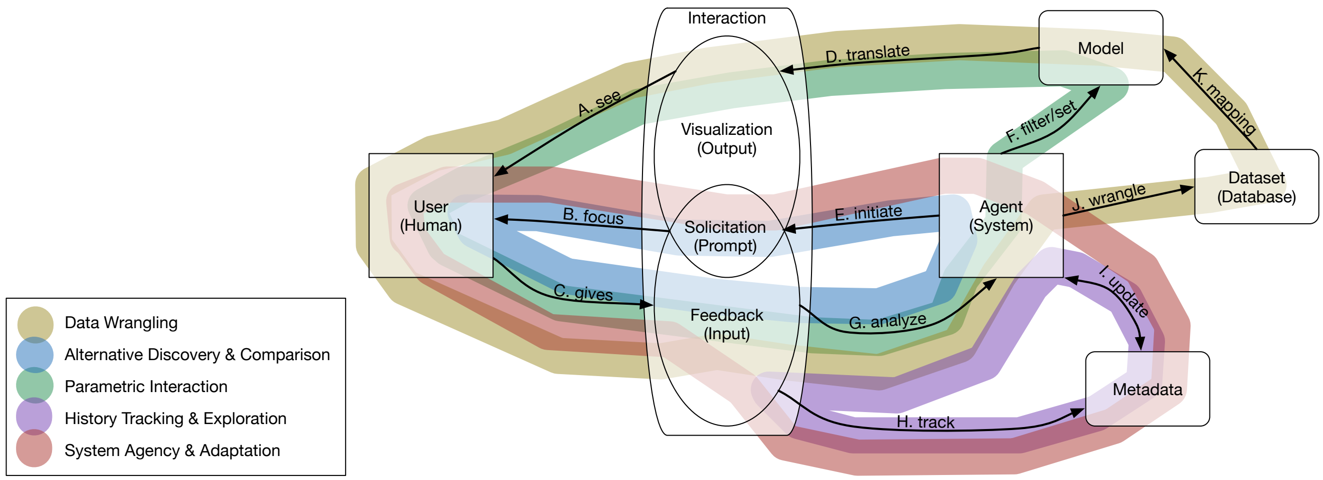


Fig. 2. The different feedback loops for each functional piece of a fully integrated MIVAS. One key aspect is that within each feedback look the *agent* may or may not be intelligent. For example, agent can represent a script or an algorithm that is run after the system receives input from the user. However, in MIVASs the agent must be intelligent enough to adapt and respond appropriately to the user's actions and intentions. Within interaction, we see solicitation as the intersection between visualization and feedback. Solicitation uses visualization techniques to obtain required feedback from the user.

interaction (Section VI), history tracking and exploration (Section VII), and system agency and adaptation (Section VIII). Figure 2 depicts the components of a complete MIVAS. Within this diagram, we highlight the *feedback loops* of each of these five functional components. As noted in Figure 2, the *agent* can represent a system or function using a fixed algorithm, a predetermined script, a semi-intelligent agent that responds differently based on statistical results, such as outlier detection, or a fully intelligent agents that can adapt and respond to the way the user interacts with the system. Fully intelligent agents are a critical part of a MIVAS. MI agents can proactively use other, potentially simpler, agents, such as algorithms or scripts, to discover potential problems or identify discoveries that the user may have overlooked and then prompt that user at the appropriate time.

A. Differentiation from Other Fields

1) Human-Computer Interaction: Human-Computer Interaction (HCI) is the study of the interface between the computer (or system) and a human. One major component of this field studies the effectiveness of said interaction. For example, the effectiveness of a particular visualization based on the lifestyle factors of different people [36]. In Figure 2, HCI focuses on the left to middle of the diagram; i.e., the user and interaction. In contrast, MI systems subsume HCI research and also encompass the system side. Building on this, MIVASs ask how the system can behave in a more dynamic and intelligent way to enhance the interaction between the user and the system.

2) Semi-Autonomous Systems: Semi-autonomous systems can only operate fully autonomously under certain conditions, but cannot complete a task without outside assistance in all circumstances [37]. Most of these systems are generally perceived as autonomous agents that require minimal human intervention only in certain situations. A simple example is a Roomba vacuum, which becomes trapped and must be realigned, or must be carried upstairs to complete its objective. This form of automation would not qualify as being MI, as the collaboration between agent and human is ultimately a shortcoming of the agent. Incorporating a semi-autonomous agent into a MI framework should improve and encourage collaboration, as the human can then contribute to the agent completing a task, or vice versa.

B. Issues in MIVAS Implementation

There are a range of concerns in properly implementing a successful MI architecture, particularly software designs intended to bring human interaction into the process [38]. In order to effectively benefit from interaction, the software must not only provide better output than fully automated or manual counterparts, but must also provide a flexible and productive work environment for the human that enhances the work performed by the user, as opposed to unintentionally slowing or impeding natural human work flow.

Issues to be addressed include identifying how and under which circumstances initiative is to be passed between the human and machine agent, how to maintain awareness within the machine agent as the exploratory task evolves over time, and what degree and means of communication exists between the human and machine agent. These concerns need to be addressed while respecting the attention, patience, and existing workload of the user. Otherwise the predictive agent becomes no more useful to the collaboration than an uninformed and/or controlling work partner. Also, greater degrees of agent-based automation must take into account the possibility of the automation failing, which might drastically reduce the human's work performance and overall confidence in the machine agent [39]. Ultimately, the machine-agent does not exist merely to assist a user's task, but also to help develop confidence in the data analytic process itself. Incorporating meaningful human interaction into the knowledge discovery process, and also the machine-agent as a more transparent and accessible collaborator, enables the user to better understand visual outputs and insights derived by the analytic software.

IV. DATA WRANGLING

While an effective visualization may significantly enhance the speed and certainty with which an analyst performs their work, the issue of data wrangling, i.e., cleaning and transforming data into a form that is suitable for analysis, is still the most tedious and labour-intensive aspect of data analytics [40]. Data wrangling is separate from data mining, which involves some method for extracting models, patterns, and general insights from data. For an analyst, any effective insight resulting from data mining must be preceded by some measure of data quality and confidence resulting from a wrangling phase. To perform large-scale data transformations, analysts often write custom scripts to clean and standardize raw data. These scripts greatly reduce the time required for data wrangling when compared to manually editing datasets. However, they require some level of programming proficiency and ultimately are only as intelligent as the analyst writing them. Here we list data wrangling software that builds on collaboration between users and machine agents.

Figure 2 shows feedback loop K-D-A-C-G-J. A *dataset* is loaded then mapped to a model. The *model* is translated to a *visualization* that the user sees. As the *user* gives *feedback* about the data, the *agent* analyses and then wrangles the data.

A. Research Examples

Here we describe two existing examples of approaches to *data wrangling*. However, we do not include a summary of this in Table I because data wrangling is generally available and works reasonably well.

1) TimeCleanser: TimeCleanser [41] is a data wrangling package specifically designed for cleaning time-series, a form of data which is sensitive to ill-formatting and unit invariance. Because TimeCleanser is specialized for time-series, collaboration between the user and the system agent is predominantly a fixed-initiative or semi-automatic process. The dataset is checked against a list of syntax-related quality measures with more complex checks for plausibility of the input. Additionally, user-definable checks can be performed once the user becomes more familiar with the dataset. Anomalies and plausibility issues are visually represented side-by-side with data tables allowing the user to quickly identify erroneous inputs and resolve them.

2) Trifacta: Extending previous research in MI data wrangling, Trifacta [42] is a commercially available system that assists analysts in cleaning and preparing data prior to the core analysis. The system intelligently assesses datasets and visually reports outliers, distributions, missing, or inconsistent values to the user. The software permits the user to quickly generate additional data fields, create annotations for entities of a dataset, and combine different datasets into one. The system infers new recommendations from the contents of a dataset towards shaping and preparing data to maximize the effectiveness of future analytic exploration. In this sense, Trifacta goes beyond standard spreadsheet manipulations, in that it employs an assertive agent that can adapt to the particular context of a dataset, intelligently guiding the user to create datasets that are optimized for potential visual exploration. Trifacta observes the users' involvement, automatically recording interactions and generating transformational scripts, which can be saved and exported for use on other datasets. This not only creates a reference of data provenance, but also the time saving benefit of using scripts without having to write them.

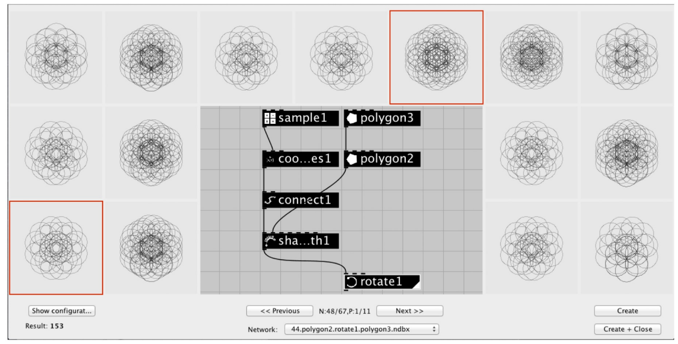


Fig. 3. An example of exploring alternatives in GEM-NI [43].

V. ALTERNATIVE DISCOVERY & COMPARISON

VA aims to use the faculties of human visual processing to richly derive clear and salient meaning from a visual representation of data. To achieve this, analyst often must repeatedly fine-tune both modelling choices and visual aesthetics, and compare alternative visualizations in search of both insight and the appropriate means of communicating that insight. This often involves an analyst jumping back and forth between alternatives, either using basic measures to manually capture possible outputs to contrast, or making adjustments to a visualization and comparing it against the memory of past alternatives. Having software that assist the user in navigating, displaying, and even creating alternatives can greatly decrease the cognitive workload placed upon the analyst and ultimately increase the quality of the output.

Figure 2 shows feedback loop E-B-C-G. The *agent* initiates *solicitation* requiring feedback on generated alternatives. The *user's* focuses on the prompt and gives *feedback* by selecting one or more of the alternatives. The agent then analyses the feedback and acts accordingly.

Recent research in generative graphic designs presents new methods for navigating and creating alternatives for generative networks; i.e., data-flow programs [43]. GEM-NI includes functionality for branching off new alternatives, side-by-side comparisons between them, parallel editing, resurrecting previous states from a visualized history, and the capability to merge progress from one branch of exploration into another (Figure 3). Further, GEM-NI introduces a structural design gallery, which generates a set of *parametrically and computationally different* alternatives for further exploration. A MI approach to game-level design has shown the effectiveness of managing possible alternatives to foster co-creativity between human and machine agents [44]. Such software attempts to bolster diagrammatical lateral thinking by rapidly generating possibilities for the user to evaluate, judge, and thus guide the user through the space of possible creative alternatives in an iterative manner.

A. Software Examples

Here we describe examples of existing research that implements *alternative discovery and comparison* (see Table I for a summary of the discussion below).

1) ViA: ViA [9] is an assistive AI-agent that produces visualizations based on hard-coded guidelines that reflect the nature of human perception/cognition, search heuristics meant to maximize visual salience, and information pulled from user interactions. To find a meaningful and visually salient visual mapping of a dataset, the user first answers a series of questions to help constrain the search, and is then presented with a gallery of possible visualization alternatives for the user to select from. The mapping search is run alongside an evaluation engine, which rates the search results against a hard-coded set of criteria informed by human visual perception. With this evaluation, the machine-agent indicates a notion of uncertainty and confidence regarding the effectiveness of visualization alternatives, and thus further assist the users' decision making.

2) LineUp: LineUp [19] is an interactive attribute ranking software, part of the Caleydo Project (see *http://caleydo.github.io*), a larger open-source suite of VA solutions. Users can perform transformational operations on attributes of entries, which immediately update the visualized list of entries. Snapshots of visual lists can be created and displayed side-by-side to help the user view how filtering on different attributes changes the rankings.

3) ExPlates: ExPlates [45] is an approach to Exploratory Data Analysis (EDA) that uses connected nodes, or Plates, to represent steps and iterations of computational functions and interactive visualizations. EDA lets the user think diagrammatically as it encourages multiple data-driven hypotheses to be investigated. However it places a high cognitive load on memory, perception, and cross hypothesis reasoning. ExPlates provides a workspace to offload part of this cognitive load by spatially representing the exploratory history through a branching trail of visualizations and their interconnected functions. While this visual branching is effective at navigating and comparing multiple alternatives, ExPlates does not offer any agent initiated assistance in discovering alternatives.

4) TimeFork: TimeFork [8] is a MI platform for analyzing predictions based on time series data, focused primarily on stock market trading. The software provides both an overview and isolated view of past stock market data. By interacting with the visual representations of stock data, the user can balance between various predictive algorithms and user generated predictions, viewing the immediate and interdependent effects of buying and selling stocks in a side-by-side display. This approach enables the analyst to leverage their domain knowledge and expertise against the results of predictive analytic algorithms to determine the best coarse of action.

5) READ: READ [7] is an EDA software, which breaks the overall process of data exploration into segments that can be defined and reused. READ aims to reduce EDA's time-intensive component of iteratively producing and evaluating visual models, as well as to reduce the likelihood of the user-driven exploration missing a viable modelling alternative. Following a data preparation phase, READ enumerates a pool of possible models for the user to search through, offering validation metrics that help to identify the more interesting models. The pool of alternatives can further be sorted via clustering algorithms, allowing the user to interact with parametric control panels to fine-tune clusters in search of interesting models. Clusters can be annotated for later analysis.

VI. Parametric Interaction

There are many ways a user may visually interact with data representations [46]. Some interactions focus on querying or highlighting visualizations, while others are intended for manipulating data mappings and aesthetic views. These latter forms of interaction involve the user manipulating predefined parameters within models, which map data to a visual representation, and often require expert domain knowledge, particularly a knowledge of how a visualization is being produced by a given mapping algorithm [47]. Such parametric interaction is typically supported by representing the set of parameters within a control panel, either as a button, an adjustable numerical value, or a slider. While this degree of representation absolves the user of directly manipulating modelling algorithms, it often results in a repetitive trial and error workflow.

Figure 2 shows feedback loop D-A-C-G-F. A *model* is translated into graphical form as a *visualization*. The *user* sees it and gives *feedback* in response. The *agent* analyses the feedback and sets new *model* parameters and may filter out data to create a new model.

A MIVAS should provide some means of assisting the user in these interactions, by creating an interface that is understandable and intuitive for human workflow, and passing off the computational and mathematical elements of data analysis to the machine agent. In this sense, the visualization becomes a common workspace between the user and the predictive agent, whereby interactions are a means of bi-directional communication between the human and agent, a feature that is fundamental to implementing an effective MI system [48]. Rather than using a natural language dialogue, which currently limits the range of understanding for the machine-agent, the use of visual metaphors can play an important role as a basis of communication through workspaces. Primary visual metaphors [49], such as quantity or importance is size, similarity is proximity, intensity is heat, or movement is time, are easily understood by human reasoning and can be artificially interpreted and generated through calculation.

While not all interactive workspaces provide a mixed collaboration between a human and an adaptive and autonomous machine agent, it is still possible to discuss a form of intelligent interaction that maximally interweaves the capacities of humans reason and judgment with the computational and storage capacities of a reactive agent. Such an agent may not develop and adapt to the evolving goals over the course of the analyst work, and as such may not initiate any unsolicited assistance. Yet, we a reactive agent can shadow the analytic initiatives of a human user, automatically assisting with a predefined set of computational and record keeping strengths.

A. Research Examples

Here we describe examples of existing research that implements *parametric interaction* (see Table I for a summary of the discussion below).

1) ForceSpire: ForceSpire [50] is a text analysis software designed as an open and interactive spatialized workspace that incorporates force-attracted graph visualizations and adaptive clustering algorithms. It permits users to layout, highlight, and annotate text documents, through a visualized representation of document relatedness through proximity and links. The ForceSpire software offers an approach for adjusting parameters by having the machine-agent interpret the user's interactions according to a spatialization metaphor of visual proximity to express semantic similarity. As a user drags, pins, annotates, and highlights text entities within the workspace, metadata is created and used to update parameters within the clustering algorithm, which in turn update the visualization. This creates an active loop between the user and the predictive system, which occurs at a semantic/visual level.

2) ScatterGather: Scatter-Gather [51] is a visual analytic platform for clustering algorithms that uses a MI approach to incorporate iterative user feedback in order to fine tune the clustering output. Users are presented with a visualization of an initial clustering, and can interactively and iteratively update constraints through simple interaction in a control panel. However, the user interacts with the model through making decisions on splitting or combining visually represented clusters, not with clustering k-values, which allows the workflow to be more centred on the analysis of the visual output than on the clustering algorithms parameters.

3) LineUp: LineUp [19] incorporates visualized parametric interaction by permitting users to witness real-time visual updates through interacting with object-level representations of parameters. While the visualized parameter control panel is not an interpreted parametric interaction, the visual representations provide a more intuitive workflow for parametric manipulation than merely adjusting numerical values.

4) BixPlorer: BixPlorer [52] is an interactive analytic workspace that uses *bi-clusters* as its main object for analyzing relations within text documents and between entities. A bi-cluster is two overlapping clusters or sets, where every entity within one set is related to every entity from the other set. By using bi-clusters as the primary interactive tool for exploration, analysts can chain different bi-clusters together to rapidly focus the scope to reveal interesting relations and connect these relations to associated text documents. The interactive workspace permits the user to spatially arrange bi-cluster and related text elements, as well as highlight and annotate important findings, to reduce the typical cognitive load of exploratory data analysis. Bi-clusters elements are also visually represented and the software can help draw attention to important entities within a bi-cluster through a gradient colour scheme that adapts to the ongoing connections the user generates while exploring.

VII. History Tracking & Exploration

A critical element in data analysis environments is the issue of data provenance. While humans are inherently well equipped at reasoning and contextualizing, it is only with an inherently limited scope of perceptual attention and available working memory [53], which makes it inherently hard to keep track of where (all) data came from. Figure 2 shows quasi feedback loop G-H-I. *Feedback* is simultaneously analyzed but the agent and tracked by recording *metadata*. Once the *agent* finishes its analysis it updates the metadata with actions taken in response to feedback received.

Data Provenance is used to track the movement and transformation of data within and between datasets as the user progresses and develops insight into the data [54]. Software that can assist the user with data provenance is useful or at times necessary in systems that focus on human exploration through manipulation of data and visualizations, as the evolutionary process of exploration can often mask how or from where certain insights are derived. Fully automated systems might skip this step altogether, providing the user with an output that must be trusted without enabling an understanding how or under what set of presuppositions it was produced. In a MI framework, data provenance is often assigned to the machine-agent as a fixed initiative task whereby human work and exploration is automatically rendered into an accessible history of data transformations or branching avenues of exploration.

The key to effective data provenance is not merely to take advantage of the superior memory capabilities of computers, but to make this history accessible and useful to the exploration process itself. This may include: allowing annotations from the user; creating an action list based on user interactions (e.g., to remind current and future users of what has occurred); automatically creating exportable scripts for future data transformations; and incorporating a visual history of exploration to enable users to quickly and informatively return to prior states of exploration without losing their progress.

A. Metadata

In addition to preserving a retractable path of exploration and history of data transformation, a system may also store information derived from user interactions with a particular dataset, such as derived insights, semantic relevance, and/or user preferences. Such secondary information, known as metadata, is stored separate from the original dataset, and can be used as contextual guidance for machine-agent assistance. While some researchers encourage the interaction between initial and secondary data to enhance visual analytic output [26], [55], exactly how and what secondary information is to be extracted from a user is still open to interpretation.

B. Software Example: CzSaw

CzSaw [56] is an analytic workspace for visualizing document collections and entity relations in a graph format (summary in Table I). The software offers a user driven workflow accompanied with automatic user interpretation, which generates an accessible, navigable, and reusable analytical work history. As the user interacts with the visual representation, filtering, connecting entities, and adjusting parameters within the model, the software translates these actions to scripts and generates dependency graphs displaying the branching history of data transformations. These scripts are easily accessible within the graph and can be reused to quickly duplicate entity modifications. A visual timeline of analytic processes enables the exploration of alternative solution paths through split editing lineages.

VIII. System Agency & Adaptation

Early research on MI designs categorized four stages or levels of interaction that a machine-agent could undertake [57]. These are *unsolicited reporting*, *sub-dialogue initiation*, *fixed subtask initiative*, and *negotiated mixed initiative*. The first two levels involve prompting or dialoguing with the user to report critical information or to clarify ambiguous information. A fixed-initiative task is a predefined task that the machine-agent attempts to solve when required, a definition, which matches most implementations of semi-automatic interactions. Yet, negotiated initiative is the truest form of mixed collaboration. It involves communication and a shared awareness of both the task at hand and the capabilities each collaborator has to offer in finding a solution. The timing of collaborative interaction is crucial to ensure success.

Figure 2 shows feedback loop C-H-I-E-B. The *user* gives *feedback*, which is tracked using *metadata*. The *agent* can receive and then update the metadata. If the agent takes initiative and makes a substantial enough discovery it may decide to *solicit* feedback from the user.

Given a specified circumstance or having met a condition of relevancy, a machine agent may interact with or engage the attention of the user by presenting a possible discovery or insight within the data to the user. This may be a clustering discovery, a particular visual mapping that highlights some observation, an unforeseen result of different data mining algorithms, or emphasizing certain avenues of exploration for the user that have been neglected or might prove fruitful. A few existing VA systems provide some of these features in a fully automated manner. However, a MI framework considers these features to be best utilized and most effective when human reasoning is kept *within* the knowledge discovery loop.

The idea of "keeping the human in the loop" still places the machine-agent at the center of the reasoning process, with the human providing guidance on the overall outcome. Following the notion that "a little domain knowledge goes a long way", Endert *et al.* [50] propose knowledge discovery loops that are centred around the analysts work environment. This encourages data exploration to be initiated primarily by human interest, while the machine-agent provides guidance/insight. To this end, the use of flexible, scalable exploratory workspaces, interfaces which encourage spatial organization and planning of data entities, is not only an effective means of visually capturing and off-loading the process of exploratory reasoning, but can become a means of situating machine-awareness to the analysts task [53].

A. Software Examples

Here we describe examples of existing research that implements *system agency and adaptation* (see Table I for a summary of the discussion below).

1) RESIN: RESIN [6] is a predictive analytics system that combines human analytic reasoning with an AI blackboard agent, a fixed-initiative task modeller, and a visually interactive analytics interface. The system focuses on the blackboard component, which is used to create and monitor multiple hypotheses. The blackboard agent assists in determining problem solutions, creating a hypothesis verification loop that assigns analytical tasks to the user, who in turn returns the results to the blackboard to update the knowledge generating loop. As the task evolves through verification and dismissal of various hypotheses, the system can make well-informed predictions, as well as provide measures of confidence for various predictions to the user. The task-structuring agent within RESIN also adapts to shifting time-constraints as the project's goals evolve and new predictions are being verified.

2) PerCon: PerCon [5] is a dataset library management tool that helps visually store and navigate large amounts of datasets. Users can spatially arrange datasets within an analytic workspace, generating implicit metadata regarding relations between datasets. The system monitors and generates further metadata derived from explicit user inputs and searches. This additional information is used to propose relevant, yet unviewed datasets to the user, based on a history of recent user interactions that reveal specific interests. The user can also directly request recommendations that indicate explicit interests.

3) ForceSpire: In the ForceSpire system [50], user interaction with text entities (arranging semantically similar documents closer together in the workspace) is observed and interpreted by the predictive-agent, which may then present hypotheses regarding the possible semantic relations between text documents to the user. Rejecting or accepting the agent's hypotheses strengthens specific semantic terms in the clustering algorithm, and the visualization is updated accordingly.

4) ALIDA: ALIDA [58] is an adaptive agent designed to determine a users interest while interacting with a VA interface. The agent autonomously monitors users interface behaviour and makes predictions after every minute based on captured interactions. The agent can decide upon an area of interest that the user is browsing, if results appear too scattered, or that the user is inactive based on little to no recent interactions. ALIDA records all past decisions as metadata for multiple purposes. First, past decisions about user interests are weighed into the decision function to increase decision accuracy. Second, the metadata becomes an accessible history of interest, similar to editing or interaction histories, which the user can potentially explore through, functioning as a reminder to momentarily forgotten areas of interest. This metadata may also be expressed indirectly by rendering specific interests within the interface more visibly predominant than others.

IX. Discussion

MIVASs are inherently complex and we need to understand the computational and implementation aspects of such systems. MIVASs needs to be able to differentiate between exploratory analysis and focused analysis tasks. The user's task type should modify the interaction style between the user and the agent. For example, in an exploratory analysis the agent might run a long term analysis that is not time critical as opposed to a focused analysis task, where responsiveness is more important and the agents need to adapt more quickly to user goals.

There are a number of areas that can benefit from a MI approach, from human-human collaboration to methods of communication between users of varying technical abilities, such as a group of non-technical people, analysts, engineers, and programmers. Based on the type of user, the MI agent should prompt at different levels of severity if a problem is found. Prompts should reflect user appropriate details and wording including appropriate responses. For example, a simplistic response or prompt could frustrate an expert user, while a complex message may overwhelm a non-technical person.

Agent assisted human-human collaboration is a topic in which MI approaches will pay a key role [59]. This includes multi-user user collaboration where users who are not in the same location or cannot work at the same time, e.g., due to schedules or time zones, work together. For remote collaboration, MI agents can focus attention on important changes made by other remote collaborators. This also includes the synchronization of data to achieve the responsiveness of a system when coordinating distant interactions. For collaboration based on availability, MI agents can support a collaborator by showing what other collaborators have done recently. The agent could recommend or notify the collaborator about whether the changes made by others fit with the goals and intentions of others or cause conflict.

Another approach is that the system autonomously revisits the work of the user and checks if there is a more optimal solution. For example, if the user used a data-mining algorithm, the system could try other such algorithms in the background and prompt the user if it finds one that yields both better precision and recall for the current dataset. One interesting user interface aspect of this approach is that such reports from the system are likely best presented as alternatives to the user; i.e., an alternative solution that does not override the users work. This then enables the user to compare and contrast the systems' work with their own work and to analyze if the new solution is truly better for the users' goal.

Key concerns and considerations that need further investigation are *how does a MI system deal with uncertainty in the human's decision making process* and *how do we insure that the agent does not introduce more uncertainty into the system?*. If not addressed properly, these concerns can make any system unusable. Further, *when do agents take initiative? And, how much initiative should they take?* The system must also be able to identify the degree of erratic human interaction. Erratic human interaction can come in the form of frustration, lack of a focused used of the systems, or just exploration (ordered for negative to positive). When a negative interaction is encountered the system agent should solicit and visualize differently than with more positive interactions. At this point we can only say that this is an issue that needs to be accounted for in any MI system. Further, we can say that agents need to be able to learn and anticipate human interactions (which can be different for each user of the system) and if an agent's initiative is inappropriate then the agent needs to back off and adapt, or unlearn that part of the anticipation.

X. Conclusions

In this survey we have discussed the importance of MI research. We have shown that MI research is more then VA. MI research expands into other key areas where the system interacts intelligently with the user, negotiating who leads analytical discovery – sometimes the human, sometimes the computer. This is done through predictive machine learning where the system learns to anticipate what the user wants and proactively accomplishes this task for the user.

Our contribution is the description of our conceptual architecture of a mixed-initiative visual analytics system (MIVAS). A MIVAS has five key components: data wrangling, alternative discovery and comparison, parametric interaction, history tracking and exploration, and system agency and adaptation (an additional contribution). As discussed, these components appear separately in previous work, but not collectively in a single system. For any system to be considered to be a full MIVAS the missing components from Table I need to be added..

Acknowledgments

This work was supported by the NSERC CRD grant 461882-2013. Thanks to our industry collaborator Fuseforward Solutions Group Ltd.

References

[1] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *Computer Graphics and Applications, IEEE*, vol. 26, no. 1, pp. 10–13, 2006.

[2] J. MacInnes, S. Santosa, and W. Wright, "Visual classification: Expert knowledge guides machine learning," *Computer Graphics and Applications, IEEE*, vol. 30, no. 1, pp. 8–14, 2010.

[3] E. Bertini and D. Lalanne, "Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 9–18, 2010.

[4] ——, "Surveying the complementary role of automatic data analysis and visualization in knowledge discovery," in *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration*. ACM, 2009, pp. 12–20.

[5] S. Park and F. Shipman, "Percon: A personal digital library for heterogeneous data," in *Digital Libraries (JCDL), 2014 IEEE/ACM Joint Conference on*, Sept 2014, pp. 97–106.

TABLE I
COMPARING CURRENT RESEARCH WITH THE NECESSARY FUNCTIONALITY OF MIXED-INITIATIVE

Research	**Alternative Discovery & Comparison**	**Parametric Interaction**	**History Tracking & Exploration**	**System Agency & Adaptation**
ALIDA [58]		interactions are recorded to discern user interests	records agent's decisions as interactive metadata	past agent decisions allow more probable discernment in tracking user's interests
BixPlorer [52]		exploratory chaining of biclusters; colour communicates important entities	user annotation; spatial map	
CzSaw [56]		interactions are translated to scripts to generate transformational dependancy graphs	visual history of analysis process; dependancy graph allows the reuse of scripts	reactive agent
ExPlates [45]	branching exploration viewed in workspace	node/tablet operations	user annotation; spatial map/exploratory history	
ForceSpire [50]		graph spatialization translated to clustering algorithm	user annotation; spatial map/exploratory history	cluster retraining; creates user soft data; hypothesis dialog
LineUp [19]	side-by-side comparison of rankings based on attribute changes	interaction with visual parameter representations to adjust model parameters		
PerCon [5]	recommends relevant unviewed datasets		user annotation; spatial map; metadata on a dataset on relations between datasets, and user preferences	agent incorporates user searches and recorded metadata to recommend relevant datasets
READ [7]	rapidly enumerates search space of models		metadata from data preparation accessible to user	dialogue used to focus generated models
Resin [6]	blackboard-agent can help identify alternate solution paths	users interact with problem solving agent to set goals		blackboard-agent negotiates hypotheses, evidence, and predictions; task agent creates time sensitive schedule
ScatterGather [51]		control panel interaction to split and combine representations of clusters		
TimeFork [8]	side-by-side comparison of multiple predictive algorithms; can chain these to create what if paths	user interactions generate cascading predictions		predictive agent collaborates by incorporates user generated hypotheses
Trifacta [42]		user interactions create transformation scripts	can record scripts and export to other datasets	recommendations for data use based on data types
Via [9]	mapping options presented to user	parameter initializing questions; user feedback used to focus generated models		hints offered to improve mapping quality; expected utility of hint calculated

[6] J. Yue, A. Raja, and W. Ribarsky, "Predictive analytics using a blackboard-based reasoning agent," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, vol. 2. IEEE, 2010, pp. 97–100.

[7] U. Khurana, S. Parthasarathy, and D. S. Turaga, "Read: Rapid data exploration, analysis and discovery." in *EDBT*, 2014, pp. 612–615.

[8] S. K. Badam, J. Zhao, N. Elmqvist, and D. S. Ebert, "Timefork: Mixed-initiative time-series prediction," in *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*. IEEE, 2014, pp. 223–224.

[9] C. G. Healey, S. Kocherlakota, V. Rao, R. Mehta, and R. St Amant, "Visual perception and mixed-initiative interaction for assisted visualization design," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 2, pp. 396–411, 2008.

[10] J. Kohlhammer, D. Keim, M. Pohl, G. Santucci, and G. Andrienko, "Solving problems with visual analytics," *Procedia Computer Science*, vol. 7, pp. 117–120, 2011.

[11] D. Sacha, A. Stoffel, F. Stoffel, B. Kwon, G. Ellis, and D. Keim, "Knowledge generation model for visual analytics," 2014.

[12] D. Fisher, S. M. Drucker, and A. C. König, "Exploratory visualization involving incremental, approximate database queries and uncertainty," *IEEE computer graphics and applications*, no. 4, pp. 55–62, 2012.

[13] R. M. Edsall, M. Harrower, and J. L. Mennis, "Tools for visualizing properties of spatial and temporal periodicity in geographic data," *Computers & Geosciences*, vol. 26, no. 1, pp. 109–118, 2000.

[14] W.-H. Hsu, J. Mei, C. D. Correa, and K.-L. Ma, "Depicting time evolving flow with illustrative visualization techniques," in *Arts and Technology*. Springer, 2010, pp. 136–147.

[15] J. Woodring and H.-W. Shen, "Chronovolumes: a direct rendering technique for visualizing time-varying data," in *Proceedings of the 2003 Eurographics/IEEE TVCG Workshop on Volume graphics*. ACM, 2003, pp. 27–34.

[16] M. Krstajic, E. Bertini, and D. A. Keim, "Cloudlines: Compact display of event episodes in multiple time-series," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2432–2439, 2011.

[17] B. Alper, N. H. Riche, G. Ramos, and M. Czerwinski, "Design study of linesets, a novel set visualization technique," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2259–2267, 2011.

[18] C. Shi, W. Cui, S. Liu, P. Xu, W. Chen, and H. Qu, "Rankexplorer: Visualization of ranking changes in large time series data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2669–2678, 2012.

[19] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit, "Lineup: Visual analysis of multi-attribute rankings," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2277–2286, 2013.

[20] L. Zhang, A. Stoffel, M. Behrisch, S. Mittelstadt, T. Schreck, R. Pompl, S. Weber, H. Last, and D. Keim, "Visual analytics for the big data era – a comparative review of state-of-the-art commercial systems," in *Visual*

Analytics Science and Technology (VAST), 2012 IEEE Conference on. IEEE, 2012, pp. 173–182.

[21] J. R. Harger and P. J. Crossno, "Comparison of open-source visual analytics toolkits," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 82 940E–82 940E.

[22] S. S. Salim *et al.*, "A systematic review of shared visualisation to achieve common ground," *Journal of Visual Languages & Computing*, vol. 28, pp. 83–99, 2015.

[23] G.-D. Sun, Y.-C. Wu, R.-H. Liang, and S.-X. Liu, "A survey of visual analytics techniques and applications: State-of-the-art research and future challenges," *Journal of Computer Science and Technology*, vol. 28, no. 5, pp. 852–867, 2013.

[24] J. Lapa, J. Bernardino, and A. Figueiredo, "A comparative analysis of open source business intelligence platforms," in *Proceedings of the International Conference on Information Systems and Design of Communication*. ACM, 2014, pp. 86–92.

[25] F. Serban, J. Vanschoren, J.-U. Kietz, and A. Bernstein, "A survey of intelligent assistants for data analysis," *ACM Computing Surveys (CSUR)*, vol. 45, no. 3, p. 31, 2013.

[26] D. L. Hall, J. Graham, and E. Catherman, "A survey of tools and resources for the next generation analyst," in *SPIE Sensing Technology+ Applications*. International Society for Optics and Photonics, 2015, pp. 94 990B–94 990B.

[27] N. Mishra and S. Silakari, "Predictive analytics: A survey, trends, applications, oppurtunities & challenges," *International Journal of Computer Science and Information Technologies*, vol. 3, no. 3, pp. 4434–4438, 2012.

[28] P. D. Yoo, M. H. Kim, and T. Jan, "Machine learning techniques and use of event information for stock market prediction: A survey and evaluation," in *Computational Intelligence for Modelling, Control and Automation, International Conference on*, 2005, pp. 835–841.

[29] R. C. Prati, G. Batista, and M. C. Monard, "A survey on graphical methods for classification predictive performance evaluation," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 23, no. 11, pp. 1601–1618, 2011.

[30] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: a survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 4, pp. 421–436, 2012.

[31] E. Horvitz, "Uncertainty, action, and interaction: In pursuit of mixed-initiative computing," *IEEE Intelligent Systems*, vol. 14, pp. 17–20, 1999.

[32] W. Chen, A. Grushin, and J. H. Li, "Mixed-initiative distributed simultaneous (MDS) planning," in *Collaboration Technologies and Systems (CTS), 2012 International Conference on*, May 2012, pp. 84–90.

[33] M. Ai-Chang, J. Bresina, L. Charest, A. Jonsson, J. Hsu, B. Kanefsky, P. Maldague, P. Morris, K. Rajan, and J. Yglesias, *MAPGEN: mixed initiative planning and scheduling for the Mars' 03 MER mission*. Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2003.

[34] T. M. Green, W. Ribarsky, and B. Fisher, "Building and applying a human cognition model for visual analytics," *Information visualization*, vol. 8, no. 1, pp. 1–13, 2009.

[35] R. J. Crouser and R. Chang, "An affordance-based framework for human computation and human-computer collaboration," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2859–2868, 2012.

[36] S. Makonin, M. Kashani, and L. Bartram, "The Affect of Lifestyle Factors on Eco-Visualization Design," in *Proceedings of Computer Graphics International (CGI)*, 2012.

[37] S. Zilberstein, "Building strong semi-autonomous systems," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 2015, pp. 4088–4092.

[38] G. Tecuci, M. Boicu, and M. T. Cox, "Seven aspects of mixed-initiative reasoning: An introduction to this special issue on mixed-initiative assistants," *AI Magazine*, vol. 28, no. 2, p. 11, 2007.

[39] C. D. Wickens, H. Li, A. Santamaria, A. Sebok, and N. B. Sarter, "Stages and levels of automation: An integrated meta-analysis," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54, no. 4. SAGE Publications, 2010, pp. 389–393.

[40] P. J. Guo, S. Kandel, J. M. Hellerstein, and J. Heer, "Proactive wrangling: Mixed-initiative end-user programming of data transformation scripts," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2011, pp. 65–74.

[41] T. Gschwandtner, W. Aigner, S. Miksch, J. Gärtner, S. Kriglstein, M. Pohl, and N. Suchy, "Timecleanser: A visual analytics approach for data cleansing of time-oriented data," in *Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business*. ACM, 2014, p. 18.

[42] J. Heer, J. M. Hellerstein, and S. Kandel, "Predictive interaction for data transformation," in *7th Biennial Conference on Innovative Data System Research, CIDR*, 2015.

[43] L. Zaman, W. Stuerzlinger, C. Neugebauer, R. Woodbury, M. Elkhaldi, N. Shireen, and M. Terry, "Gem-ni: A system for creating and managing alternatives in generative design," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 1201–1210.

[44] G. N. Yannakakis, A. Liapis, and C. Alexopoulos, "Mixed-initiative co-creativity," in *Proceedings of the 9th Conference on the Foundations of Digital Games*, 2014.

[45] W. Javed and N. Elmqvist, "Explates: Spatializing interactive analysis to scaffold visual exploration," *Computer Graphics Forum*, vol. 32, no. 3pt4, pp. 441–450, 2013. [Online]. Available: http://dx.doi.org/10.1111/cgf.12131

[46] J. Heer and B. Shneiderman, "Interactive dynamics for visual analysis," *Queue*, vol. 10, no. 2, p. 30, 2012.

[47] X. Hu, L. Bradel, D. Maiti, L. House, C. North, and S. Leman, "Semantics of directly manipulating spatializations," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2052–2059, 2013.

[48] A. Endert, S. Szymczak, D. Gunning, and J. Gersh, "Modeling in big data environments," in *Proceedings of the 2014 Workshop on Human Centered Big Data Research*. ACM, 2014, p. 56.

[49] J. Cybulski, S. Keller, and D. Saundage, "Metaphors in interactive visual analytics," in *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*. ACM, 2014, p. 212.

[50] A. Endert, M. S. Hossain, N. Ramakrishnan, C. North, P. Fiaux, and C. Andrews, "The human is the loop: new directions for visual analytics," *Journal of intelligent information systems*, vol. 43, no. 3, pp. 411–435, 2014.

[51] M. S. Hossain, P. K. R. Ojili, C. Grimm, R. Müller, L. T. Watson, and N. Ramakrishnan, "Scatter/gather clustering: Flexibly incorporating user feedback to steer clustering results," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2829–2838, 2012.

[52] M. Sun, L. Bradel, C. L. North, and N. Ramakrishnan, "The role of interactive biclusters in sensemaking," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2014, pp. 1559–1562.

[53] A. Endert, "Semantic interaction for visual analytics: Inferring analytical reasoning for model steering," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2012.

[54] P. Buneman, S. Khanna, and W.-C. Tan, "Data provenance: Some basic issues," in *FST TCS 2000: Foundations of software technology and theoretical computer science*. Springer, 2000, pp. 87–93.

[55] A. Endert, L. Bradel, and C. North, "Beyond control panels: Direct manipulation for visual analytics," *Computer Graphics and Applications, IEEE*, vol. 33, no. 4, pp. 6–13, 2013.

[56] V. Chen, D. Dunsmuir, S. Alimadadi, E. Lee, J. Guenther, J. Dill, C. Qian, C. D. Shaw, M. Stone, and R. Woodbury, "Model based interactive analysis of interwoven, imprecise narratives: Vast 2010 mini challenge 1 award: Outstanding interaction model," in *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*. IEEE, 2010, pp. 275–276.

[57] J. E. Allen, "Mixed-Initiative Interaction," *Intelligent Systems and their Applications, IEEE*, vol. 14, no. 5, pp. 14–23, 1999.

[58] T. Green, R. Maciejewski, and S. DiPaola, "Alida: Using machine learning for intent discernment in visual analytics interfaces," in *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, Oct 2010, pp. 223–224.

[59] P. Keel, N. Cooke, and M. Sither, "Improving cognitive interaction through computational collaboration agents," *Theoretical Issues in Ergonomics Science*, vol. 15, no. 5, pp. 415–431, 2014.