Knowledge Generation Model for Visual Analytics

Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, *Member, IEEE*,
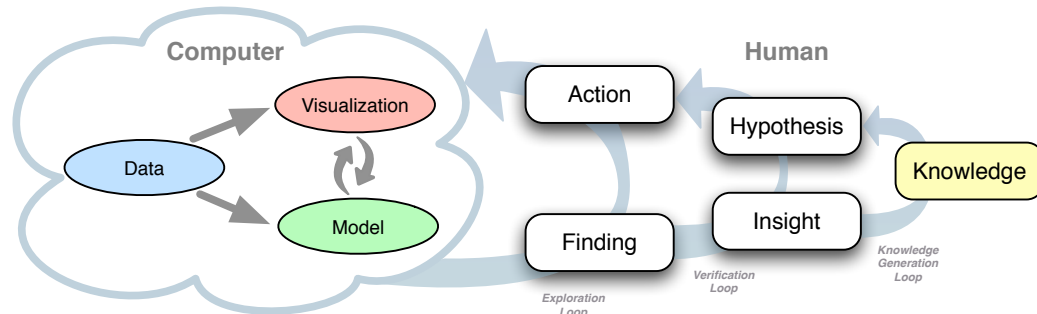Geoffrey Ellis and Daniel A. Keim, *Member, IEEE*

Fig. 1: Knowledge generation model for visual analytics: The model consists of computer and human parts. The left hand side illustrates a visual analytics system, whereas the right hand side illustrates the knowledge generation process of the human. The latter is a reasoning process composed of exploration, verification, and knowledge generation loops. Visual analytics pursues a tight integration of human and machine by enabling the user to interact with the system. These interactions are illustrated in Figure 2.

Abstract—Visual analytics enables us to analyze huge information spaces in order to support complex decision making and data exploration. Humans play a central role in generating knowledge from the snippets of evidence emerging from visual data analysis. Although prior research provides frameworks that generalize this process, their scope is often narrowly focused so they do not encompass different perspectives at different levels. This paper proposes a knowledge generation model for visual analytics that ties together these diverse frameworks, yet retains previously developed models (e.g., KDD process) to describe individual segments of the overall visual analytic processes. To test its utility, a real world visual analytics system is compared against the model, demonstrating that the knowledge generation process model provides a useful guideline when developing and evaluating such systems. The model is used to effectively compare different data analysis systems. Furthermore, the model provides a common language and description of visual analytic processes, which can be used for communication between researchers. At the end, our model reflects areas of research that future researchers can embark on.

Index Terms—Visual Analytics, Knowledge Generation, Reasoning, Visualization Taxonomies and Models, Interaction

---

✦

---

1 INTRODUCTION

Visual analytics research made great strides in the past decade with numerous studies demonstrating successes in helping domain experts explore large and complex data sets. The power of visual analytics comes from effective delegation of perceptive skills, cognitive reasoning and domain knowledge on the human side and computing and data storage capability on the machine side, and their effective coupling via visual representations. Thus far, many application papers have tested and verified different ways to instigate this human and machine collaboration to reveal hidden nuggets of information.

Recent work emphasizes that visual analytics theories must go beyond "human in the loop" to "human is the loop" thinking in order to recognize and integrate human work processes with analytics (Endert et al. [7]). To achieve this goal, system, human, cognition and reasoning based theories have to be considered. Many theoretical works have also examined the role of visual analytics tools in data analysis, decision making and problem solving. Visual analytics processes span from humans high-level analytic works using their domain knowledge to low-level activities such as interacting with tools. Many prior works

investigated different levels of activities with regards to humans cognition models (e.g., Green et al. [14], Kwon et al. [17]); interaction models/taxonomies (e.g., Brehmer and Munzner [4], Norman [25]); process models (e.g., Card et al. [5], Fayyad et al. [8]); sensemaking models (e.g., Pirolli and Card [29]). We initially sought interaction taxonomies that describe the aforementioned models. However, we lack an overarching pipeline that connects all the dots. Previously developed models (e.g., Keim et al. [15, 16]) are system-driven and are not describing in detail the human reasoning part in the visual analytics process. In particular, we have little idea how analytical components support knowledge generation processes and how analysts' intents drive the insight collection action forward. It would be valuable for future research to have an integrated framework of all processes and models relevant for knowledge generation with visual analytics.

The paper aims to take the first step to establish the knowledge generation model for visual analytics. First, we build our initial process model by significantly extending the previous models [15, 16] (Section 2). Comparing with previously developing models and theories, we show how our model fits various kinds of models (Section 3). Using our model, we find areas that existing visual analytics tools can improve (Section 4). We discuss our model showing open issues, model implications and future work (Section 5).

Our contributions can be described as follows. The model presented in this paper provides a high-level description of the human and computer processes within visual analytics systems which facilitates an understanding of the functionality and interaction between the components. This can be used in the design of new visualisation applications

---

- *Data Analysis and Visualization Group, University of Konstanz. E-mail:*
  forename.lastname@uni-konstanz.de

or in the evaluation of existing ones in terms of how their subcomponents support humans reasoning, decision making, and knowledge generation processes.

2 KNOWLEDGE GENERATION MODEL

Visual analytics uses data to draw conclusions about the world, a process, or an application field. It is a structured reasoning process that allows analysts to find evidence in data and gain insights into the problem domain. Our model of the knowledge generation process is based on the visual analytics process of Keim et al. [15, 16] and describes how knowledge is generated with this process. Pohl et al. [30] also discussed relevant theories that should be considered, besides a computer/system based view of the analytical process. They show that visual analytics system components, analytical procedures and human perceptual and cognitive activities and especially the interplay between these elements, lead to a complex process. The theories, namely sensemaking, Gestalt theories (describing problem solving), distributed cognition (interplay between humans and artifacts), graph comprehension theories (derive meaning from graphs/process visual information) and skill-rule-knowledge models (three-fold hierarchy of processing levels) are highlighted as important aspects of the visual analytics process. In the following sections, we define and relate common elements of the aforementioned models/fields/concepts in order to arrive at a better understanding of visual analytics in terms of computing and human processes.

Our visual analytics model (see Figure 1) is split into two parts. The computer system with data, visualization and analytic models, and the human component modeling the cognitive processes associated with an analytical session. The cloud in the model indicates that there is no clear separation between the computer and human part, as both parts are required for data analysis. Computers miss the creativity of human analysis that allows them to create surprising, often subtle or hidden connections between data and the problem domain. Humans are not able to deal efficient and effectively with large amount of data. In visual analytics the connection between the human and computer uses the humans interaction abilities and perception.

Knowledge generation in visual analytics comprises of abductive, deductive, and inductive reasoning processes. For a detailed definition and discussion of these reasoning processes see Peirce [27] and Magnani [21]. Abductive reasoning processes formulates hypotheses from observations that are unexpected or cannot be explained with existing knowledge. Assuming that these hypotheses are true, expectations of effects, patterns, or relations in the analyzed data are deduced. Through interactions with visual analytics systems, analysts try to find evidence and detect patterns in data to verify or falsify the hypotheses, which is an inductive reasoning process. We decided to model human cognitive processes with loops, because analysis does not follow deterministic rules but is rather chaotic or spontaneous nature. Analysts are often working on different hypotheses, tasks, or findings and can consequently be working on several loops in parallel.

2.1 Computer

2.1.1 Data

The starting point of all visual analytic systems is data. Data describes facts [8] in structured, semi-, or unstructured manner. It must be representative and related to the analytical problem, otherwise, the analytical process is unlikely to reveal meaningful relationships in the problem domain. In the visual analytics process, data creation, gathering and selection processes often determines the quality of the data. During an analysis, additional data can be created by automatic methods (e.g., clustering or classification) or by manual annotations. Hence, provenience information about data containing details about creation, gathering, selection, and preprocessing is important to estimate the trustworthiness of analysis results (see also [31]). The term metadata describes second order data or "data about data". The usage of this term is ambiguous depending on the domain and interests of users. In addition to provenience data, metadata describing the structure of data is usually handled by visual analytics system to access and display data appropriately. This sort of metadata is usually not subject to analysis as it describes

mainly data formats, which is necessity of any kind of data. The term metadata is also often used for data describing or summarizing other data, e.g., keywords or topics for documents or images. In the scope of data analysis this descriptive metadata is treated similar to data and we see no benefit or reason to treat this data differently, therefore the term data also includes this sort of metadata.

2.1.2 Model

Models can be as simple as descriptive statistics describing a property of a subset of the data or as complex as a data mining algorithm. The KDD process leads to models from data (see Figure 3). These models serve different purposes in the visual analytics process. Simple analysis tasks might be solved by calculating a single number that confirms or rejects a preconceived notion/expectation. For instance, a statistical test can lead to a conclusion whether or not to trust a hypothesis. Complex patterns or abstractions found by data mining methods can be used in visualizations by showing, for instance, clustering or classification results visually. In addition, the automatically created model can be analyzed or visualized to derive insights. For example, logistic regression models learn weights of features, which can be used to identify most important features or feature combination.

2.1.3 Visualization

A different path from data to knowledge is the information visualization pipeline using data visualizations (see Figure 3). Visualizations use data or models generated from the data and enable analysts to detect relationships in the data. In visual analytics, visualizations are often based on automatic models, for instance, clustering models are used to group data visually. Also, a model itself can be visualized, for instance, a box plot shows the data distribution of a dimension. Visualization methods for a model may vary depending on the state of the visualization. For example, in semantic zooming a visualization might use different properties of a model depending on the zoom level. Visualization is often used as the primary interface between analysts and visual analytics systems whereas understanding the model often requires more cognitive efforts.

2.2 Exploration Loop

The exploration loop describes how analysts interact with a visual analytics system in order to generate new visualizations or models and analyze the data. Analysts explore data supported with the visual analytic system by interactions and observing the feedback. Actions taken in the exploration loops are dependent on findings or a concrete analytical goal. In case a concrete analysis goal is missing, the exploration loop becomes a search for findings, which may lead to new analytical goals. Even when the exploration loop is controlled by an analysis goal, the resulting findings are not necessarily related to it but can lead to insights solving different tasks or opening new analytical directions.

2.2.1 Action

Different meanings of actions are present in the InfoVis community as they may be defined with different granularities or with differing relations. Pike et al. [28] illustrate that actions may concern user goals and tasks on the one hand and interactive visualizations on the other hand. According to recent interaction taxonomies (e.g., Brehmer and Munzner [4]), simple interactions, *How*, are related to higher level concepts, *Why*.

In our definition, actions refer to individual tasks that generate tangible, unique responses from the visual analytics system. For instance, a task can be to visualize a particular data property or calculate a model of a relationship in the data. Actions derived from hypotheses are usually complex actions, for instance, use a specific visualization method that has the potential to show interesting data. Actions derived from findings are normally simple actions, such as changing the visual mapping of a visualization or selecting a different parameter for model building. In visual analytics, analysts freely choose between visualization and modeling or a combination of both for their actions. Actions naturally provoke interactions of analysts with visual analytics systems.
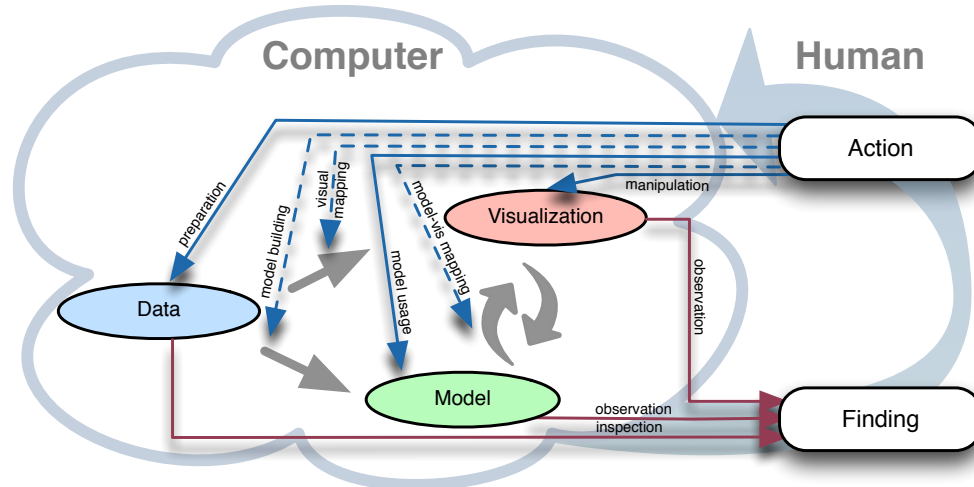
Fig. 2: Detailed part of the process model including action and cognition paths. Actions can either lead directly to visual analytic components (blue arrows) or to their mappings (blue, dashed arrows). Humans can observe reactions of the system (red arrows) in order to generate findings.

We name actions dealing with data gathering or data selection as *preparation* action because these actions are used to prepare data for the visual analytics process (Figure 2). Actions taken to create models are summarized as *model building* actions which in turn are related to the KDD process and its configuration. Application of a model is termed as *model usage*, which refers to actions such as calculating a statistic or cluster data. *Visual mapping* actions are used to create data visualizations, and *model-vis mapping* actions are those which map models into visualizations. *Manipulation* of a visualization changes the viewport or highlights interesting data in the visualization without changing the mapping of data to the visualization. All these actions are observable as interactions between analysts and a visual analytics system.

2.2.2 Finding

A finding is an interesting observation made by an analyst using the visual analytics system. The finding leads to further interaction with the system or to new insights. For example, findings from data inspection can be missing values or other data properties affecting the further analysis and require special data processing. In the case of visualizations or models, a finding can be a pattern, a conspicuous model result, or an unusual behavior of the system. Bertini et al. state that "a pattern is made of recurring events or objects that repeat in a predictable manner. The most basic patterns are based on repetition and periodicity." [2, p. 13] Patterns in data can be detected with automatic analytical methods or humans may detect patterns using their visual perception and cognition skills. Findings are in principle not limited to data, visualizations, or models but comprise of anything interesting to an analyst, e.g., an unexpected high number or a word or statement in some text.

In our definition, a finding is independent from the problem domain, however to make an analytical use of a finding, the analyst has to interpret them in the context of the problem domain. Although findings are usually associated with detecting a visual pattern, the lack of a pattern, when expected by an analyst, is also considered a finding. As shown in Figure 1, findings do not necessarily lead to new insights (see Section 2.3.2) but may trigger basic actions, such as manipulating the viewport of a visualization to show a region in more detail.

Analysts come across findings by observing the feedback from the system or examining visualizations and models, which in turn can lead to further actions. The exploration loop can be characterized as a searching activity by using the system to reveal useful findings to solve an analysis task. Analysts might frequently change their exploring strategies and switch between models and visualizations to collect different findings. The strategies and analysis directions in

the exploration loop are guided by an exploration goal defined in the verification loop. The actions and findings in the exploration loop are closely related to the visual analytics systems. Analysts gain a new insight when they are able to understand the findings and are able to interpret them in the context of the problem domain.

2.3 Verification Loop

The verification loop guides the exploration loop to confirm hypotheses or form new ones. To verify a concrete hypothesis, a confirmatory analysis is conducted and the exploration loop is steered to reveal findings that verify or falsify the hypothesis. Analysts gain insights when they can interpret findings from the exploration loop in the context of the problem domain. Insights may lead to new hypotheses that require further investigation. Analysts gain additional knowledge when they assess one or more trustworthy insights.

Findings during exploration my contribute to the verification of a concrete hypothesis but insights emerging from exploration may not be related to the examined hypothesis. It is often the case that new insights solve different analysis questions or open up new ones.

2.3.1 Hypothesis

Hypotheses play a central role in the visual analytics process. An hypothesis formulates an assumption about the problem domain that is subject to analysis. Analysts try to find evidence that supports or contradicts hypotheses in order to gain knowledge from data. In this respect the visual analytics process is guided by hypotheses. Concrete hypotheses can be tested with statistical tests or data visualizations when the data contains the necessary information. Unfortunately, hypotheses are often vague, such as the assumption that there are unknown factors that have an influence on the problem domain. In such cases an exploratory analysis strategy allows analysts to come up with more concrete hypotheses that are used for further analysis.

2.3.2 Insight

In the InfoVis community, insight has a variety of definitions. Saraiya et al. define insight as "an individual observation about the data by the participant, a unit of discovery" [33, p. 2]. North [26] takes another view and lists some important characteristics of an insight such as being complex, deep, quantitative unexpected and relevant, and going beyond an individual observation of the data. Yi et al. [43] go one step further and also consider the processes that involve insights. They focus on how people gain insight in information visualization and identify four types of insight-gaining processes, however, they argue that there is no common definition of insight. Chang et al. [6] suggest two different
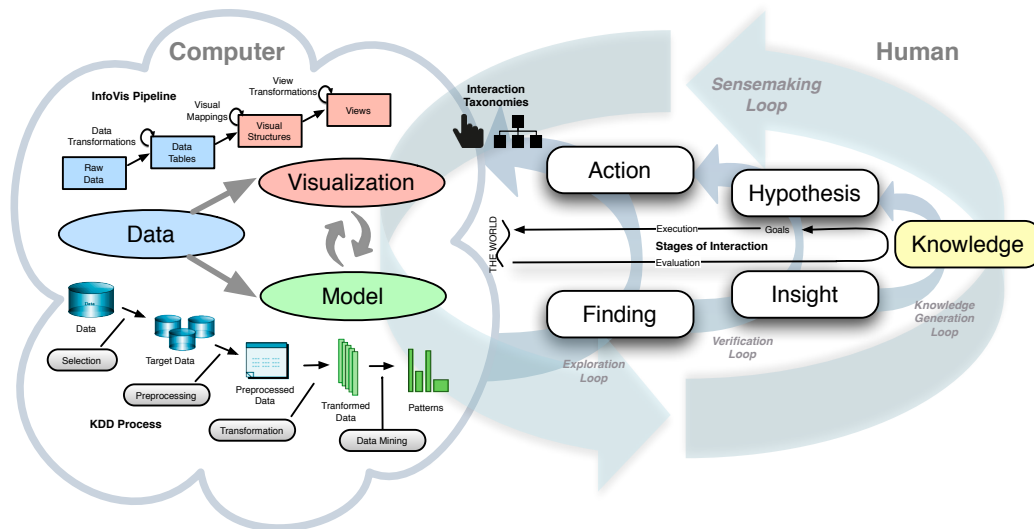
Fig. 3: Relating the process model for knowledge generation in visual analytics to other models and theories. Similarity is illustrated by color and position. A detailed illustration of interactions between the human and computer is shown in Figure 2.

definitions of insight: the cognitive science insight as a moment of enlightenment, an "Ah Ha" moment, which can occur spontaneously, and an advance in knowledge or a piece of information. Our definition is closer to the latter, where the analyst interprets a finding, often with previous domain knowledge, to generate a unit of information. Hence, an insight can be quite small, such as realizing that there is a relation between several properties of the data, to something more important and potentially significant. So, insights are different from findings in the sense that insights have an interpretation in the problem domain, what we not required for findings. For instance, a finding might support a hypothesis, which may convince the analyst and lead to the insight that the hypothesis is reliable. An analyst gains insights by collecting enough evidence to create a new hypothesis about the application or, in the case of very strong evidence, even new trusted knowledge. We consider an insight not directly as knowledge, because weak evidence might lead to an insight that needs further verification and becomes a hypothesis. For instance, finding a cluster in a visualization during an exploratory analysis might lead to the insight that there is a cluster with different properties, but this insight should at first be considered as a new assumption or hypothesis that has to be validated.

2.4 Knowledge Generation Loop

Analysts form hypotheses from their knowledge about the problem domain and gain new knowledge by formulating and verifying hypotheses during the visual analytics processes. When analysts trust the collected insights they gain new knowledge in the problem domain that may also influence the formulation of new hypotheses in the following analysis process.

2.4.1 Knowledge

Data analysis usually starts with data and one or more analysis questions. In addition, analysts bring in their knowledge about the data, the problem domain, or visual analytic tools and methodology. This prior knowledge determines the analysis strategy and procedure. During the visual analytics process analysts try to find evidence for existing assumptions or learn new knowledge about the problem domain. In general, knowledge learned in visual analytics can be defined as "justified belief" [2]. The reasoning processes in visual analytics enables analysts to gain knowledge about the problem domain from evidence found in data. The evidence has different qualities, which directly affects the trustworthiness of the concluded knowledge. The evidence collection route also impacts the trustworthiness. For example, an outcome of a statistical test of an hypothesis may be perceived more trustworthy than

a pattern found in a visualization. Depending on the collected evidence, an analyst has to decide whether enough evidence was collected to trust an insight and accept it as new knowledge or whether it is in need of further examination, e.g., analysis with different data or discussions with domain experts. Assessing the trustworthiness of new knowledge requires a critical review of the overall analysis process starting from data gathering. As well as new knowledge about the problem domain, analysts gain knowledge about data, e.g., patterns in the data or its quality. They also gain experience with the visual analytics systems and methodology.

3 RELATION TO OTHER MODELS

This section covers the most important related models that have influenced the model presented in the paper and also aims to offer a big picture of the whole knowledge generation process in visual analytics. Figure 3 illustrates related models and our knowledge generation model for visual analytics. We have used the same color codes of the original model [15, 16] to highlight related areas in our model. Feedback loops of other models (e.g. InfoVis pipeline) are replaced by our extensions on the human side. Each action results in a feedback loop via finding that may lead to new exploration loops or crosses over to higher level loops. Relevant theories are grouped in three areas according to their focus on interaction, human aspects or systems.

3.1 Systems

Card et al. propose the reference model for information visualization [5] that describes visualizations as data connected to visual mappings perceived by humans. The InfoVis-Pipeline contains the main components of *Raw Data*, *Data Tables*, *Visual Structures* and *Views* and transformations/mappings between these components which can be manipulated through *Human Interactions*.

Fayyad et al. [8] describe the Knowledge Discovery Process in Databases (KDD) as follows. KDD is the process of making sense out of data using data mining techniques at the core. The process includes the mapping of low-level data into more compact, abstract or useful forms using data mining models with the goal to discover or extract patterns that can be turned into knowledge. The KDD process consists of nine steps and represents an interactive and iterative process. Examples for interactions are selecting and filtering for the relevant data, choice of appropriate data mining methods or parameter refinement. The goal of these actions is to bring in the domain expertise of the human in order to find meaningful patterns. At each step the user might make and add decisions that cannot be handled automatically. The
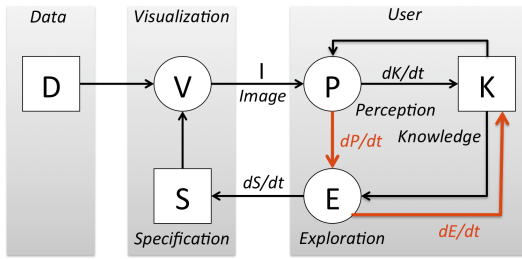
Fig. 4: Van Wijks model including Green et al.'s changes [13, 14] and labels.

KDD process steps are included in Figure 3.

As a basis for our knowledge generation model for visual analytics we take and extend the process model by Keim et al. [15, 16]. The main components of the model are *Data, Models, Visualization* and *Knowledge* each representing a stage of the process. Also a *Feedback loop* is present going from knowledge back to data. At and in between of each stage there are transitions. The visual analytics process consists of the two parts, *Visual Data Exploration* going from data via visualization to knowledge and *Automated Data Analysis* going from data via models to knowledge, as well as of their connection. Actions on the data may be performed in order to select, filter or preprocess the data. The data is then mapped to a visualization (visual mapping) and to models (data mining). Model parameters and findings are visualization and the user may interact with the visualization or refine parameters of the model in order to steer the analysis process. Knowledge can be derived from visualizations, automatic analysis and preceding interactions between visualizations, models and the human.

Figure 3 relates visual analytics components to the InfoVis and KDD system pipelines. The InfoVis pipeline corresponds to data, visualization, and the visual mappings, whereas the KDD process model is represented by data, model, and their mapping. Furthermore the visual analytics process model includes human computer interaction. Actions can be performed at each stage, namely data, visualization, model and their mappings.

The economic model of visualization by van Wijk describes contexts in which visualizations operate [40] (Figure 4). In brief, data is transferred into a visualization that can be perceived by a human through an image. Based on perception, the human generates knowledge over time which drives interactive exploration through changes to the visualization specification. Green et al. [14] add to van Wijk model, arguing that perception has an important role in interactive exploration and that the act of exploration and associated reasoning often leads to knowledge acquisition. These relate to the exploration and knowledge generation loops of our model.

3.2 Interaction

Norman describes a model for actions containing *Seven Stages of Action* [25]. At the beginning of each action each human needs a goal to be achieved. Afterwards, the human has to perform an action to manipulate something. The last step is to check if the goal was achieved. These two subprocesses are called *Execution* and *Evaluation*. Based on previous actions the action cycle is traversed several times. The model also explains two major problems that occur when interacting with computer systems. The *Gulf of Execution* indicates if humans do not know how to perform an action while the *Gulf of Evaluation* indicates that humans are not able to evaluate the result of their actions. The *Goal* concept of the stages of interaction model matches to our *Hypothesis* as a starting point. The *Execution* path is leading from *Goal* via an action to the *World*. As a result analysts evaluate observations of the *World* in several steps (see Figure 3).

Several interaction taxonomies exist which focus on different aspects, fields and domains and attempt to structure different kinds of interaction at different levels of abstraction. Most of the well-known taxonomies focus on one or two fields, e.g. visualizations (Shneiderman [36]), reasoning (Gotz et al. [12]) or data processing (Bertini et al. [2]), rather than all possible interactions in visual analytics. Recent publications attempt to structure interaction using the dimensions of *Why* (the purpose of the task), *How* (methods used to achieve this) and *What* (necessary inputs/outputs) (Brehmer and Munzner [4], Schulz et al. [34]). There are also recent taxonomies that try to integrate existing taxonomies out of different fields into a sound interaction taxonomy for visual analytics (e.g. von Landesberger et al. [41]). Taking the taxonomy of Brehmer and Munzner [4] as an example there are higher level goals that an analyst might pursue (*why*) and lower level operations that can be performed (*how*) on a specific target (*what*). High and low level interactions are both covered by the action concept. However they can be distinguished when considering the associated loops. High level actions are inspired by the verification loop and are based on insights and hypotheses whereas low level actions are influenced by the exploration loop that covers a sequence of simple actions and findings. Our process model implies that each action results in a feedback loop and is perceived and processed by the human. Our model description of action-types also shows the relevant interactions for visual analytics that are both visualization and model centric.

Also mentionable are two mantras characterizing the analysis process for information visualization and visual analytics because they shed light on human analysis strategies. Shneiderman proposes the *Visual Information Seeking Mantra* that summarizes the basic principles of many visual design guidelines [36]: *Overview first, zoom and filter, then details-on-demand.* Also Keim proposes a slightly different *Visual Analytics Mantra* [16]: *Analyse First - Show the Important - Zoom, Filter and Analyse Further - Details On Demand.* Both start with an overview/aggregation approach and end in a refinement of their hypothesis and analysis.

3.3 Human Cognition, Sensemaking and Reasoning

Humans can observe visualization or model changes that can be used for the knowledge generation process. The data may also be inspected directly.

Pirolli and Card present the Sensemaking Process [29] as a description of intelligence analysis. Central terms are the *shoebox*, *schemas* (that are a set of patterns around the important elements in the tasks), *hypotheses* and a *representation*. Sensemaking tasks are described as processes consisting of *information gathering*, the information representation as a *schema*, the generation of *insight* and finally the generation of some *knowledge product*. The first loop is called the *foraging loop* followed by the *sensemaking loop*. *Bottom-up* or *top-down*-processess are possible. That indicates that the model does not have a fixed entry point which depends on the type of task. The model shows the data and process flow with many feedback loops. Our model splits the *Sensemaking Loop* in three sub loops. In visual analytics it is also possible that the system may learn from the analysts actions allowing the system to support the user with visualization or action propositions. That is why the loop is leading through the system (see Figure 3). Hypotheses are generated as an entry point for an analysis process leading to repeated exploration cycles.

The field of reasoning and decision making both depend on the construction of mental models or scenarios of relevant situations. According to Legrenzi et al. [19] the key components of decision making are *Information Seeking*, *Making Hypotheses*, *Making Inferences*, *Weighing Advantages and Disadvantages* and *Applying Criteria to make Decisions*.

The *Human Cognition Model* (HCM) proposed by Green et al. [14] can also be applied to visual analytics. A major problem in visual analytics is that human cognition is often assumed to be an over simplified black box. Information discovery and knowledge building are at the core of the HCM. Information is presented by the computer that humans can perceive and directly interact with in order to focus their attention. The process of discovery of patterns or relations is a primary stage of knowledge that can be created within the knowledge building process. The computer works to counter the humans limited working memory as well as some cognitive biases, such as confirmation bias. Central parts of the HCM are shown in Figure 5. The HCM also includes guidelines for discovery and knowledge building. The cognition of relevant pat-
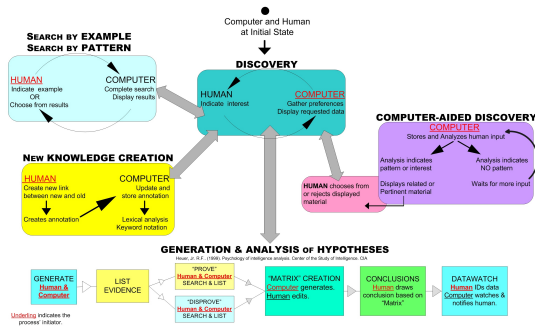
Fig. 5: Human Cognition Model by Green et al. [13, 14].

terns and schema that can be derived from insights (verification loop) and finally to some knowledge product (knowledge generation loop) plays a central role in the human part of our model. All elements of the HCM can also be found in the knowledge generation model. The key element *Discovery* describes the overall process, whereas the other sub-concepts all can be placed into our model and related to our concepts. For example, the generation and analysis of hypotheses is a central part of our model, because evidence (proof/disproof) is collected as the basis for drawing conclusions.

4 MODEL APPLICATION

In this section we illustrate that our model can be applied to systems on several levels. Firstly, the interaction possibilities can be examined according to our definition of actions in Section 2.2.1 and shown in Figure 2. Secondly, we can check if and how a system supports each of the individual loops. Thirdly, we can use our model for a comparative assessment. In the following we will demonstrate a detailed model application with Jigsaw [9, 39] and a comparative high level assessment of systems from different application domains.

4.1 Actions in Jigsaw

In the following section, we investigate Jigsaw in terms of supported actions. It is important to note that in our model each of these actions leads to its own feedback loop. Therefore, we pay special attention to system components that are capable of accepting human inputs beyond fully automated implementations. We also highlight some areas of the system that might accept more user input for amplifying user interactions.

Data preparation is not fully supported. While loading the data, the user has no possibility of adjusting any of the data preprocessing or transformation steps. It is also not possible to change those once the data has been parsed and loaded. Users may find missing data, but such manipulation must be done outside Jigsaw.

Model building is done automatically, but some views provide possibilities of adjusting the underlying models. For example, the Document Cluster View allows users to select documents as a cluster seed, which can be used when the document clusters are computed. After that, the user-generated cluster seeds become part of the model used by the application.

In terms of the *visual mapping*, Jigsaw offers some basic functionalities. In some views, users can select the background of the visualization or choose which attribute is mapped to the color of a visualization entity. Taking the color mapping as an example, there are a number of different possibilities to define such mappings, like a non-linear color scale, or the selection of the mapped colors. In this application, the authors left out any of those manipulation possibilities, and force the user to accept their pre-defined color mapping schemes.

Model usage is supported implicitly, because most of the visualizations require special models. In some views, like the Document Grid, there are different document orderings and similarity measures to choose from. This is an example of a per-visualization model usage, which can be adjusted by the user.

The *model-vis* mapping is available in almost all views, but only at a basic level. This includes various ordering, sorting and filtering options.

In general, though, this mapping is done as the developers designed it, and does not allow the user to change much.

Visualization manipulation capabilities are mostly used to highlight document instances. There are also possibilities of manipulating and changing the visual appearance of a view. For example, in the Circular Graph View, single terms can be selected, and the connections of this term to all others are displayed.

In this examination, we show that Jigsaw supports all actions we proposed in our model. It allows the visualization of different models based on the same data set, the visual mapping can be adjusted as designed by the authors, and the model-vis mapping can be modified to fit different analytical questions. In order to achieve a stronger coupling between the visualization and the underlying model, the interactions could be extended to modify the underlying model or algorithm parameters, for example when the user moves a document to another cluster in the Document Cluster View.

4.2 Knowledge Generation in Jigsaw

Having examined what actions are supported by Jigsaw, we now focus on an evaluation of the three loops, which are part of the reasoning process as defined in Section 2.

Undoubtedly, describing the human reasoning processes using a visual analytic system is complicated. This involves a variety of aspects from perception, cognition and reasoning. The detailed description and evaluation of these processes exceed the scope of this paper and requires a study with expert users. We therefore limit this examination to a description of how Jigsaw supports human reasoning.

The *exploration loop* is broadly supported by Jigsaw, by providing a number of specialized visualizations for different analytical questions. For example, users can explore given data sets, to get an impression of the contained topics, by opening the Document Cluster view which gives a labeled view of document clusters. It is also possible to modify the number of generated clusters while exploring a once-generated cluster view. The history of consecutive runs of the clustering algorithm with different parameters is stored. This allows for assessment of parameter changes with respect to topic changes. All views include a bookmark feature, which can be used to switch between different saved states of the visualization. As a result, users can capture and annotate findings that might be of further interest or of high importance for further analysis. Bookmarks can be used to store the result of different runs of the exploration loop, which in turn is done by utilizing the actions as described in the previous Section 4.1.

The *verification loop*, tightly integrated with the exploration loop, guides users to develop findings into insights. These findings from the exploration loop can be used to verify or falsify a concrete hypothesis, which is the beginning of knowledge generation. The second investigative scenario given by Görk et. al. in [9] contains some questions which can only be answered by using the verification loop. In one of the examples, Jigsaw is used to examine the sentiment of car reviews. This is done to verify the car rating, which is given as a numeric score. The sentiment, displayed in the Document Grid, is expected to agree with the score in a way, that highly rated cars have more positive reviews and those with bad ratings have more negative reviews. In this example, the positive correlation of two measures is used to verify a hypothesis, which had been inferred based on product reviews contained in the analyzed data set. Natively, Jigsaw does not support the concepts of findings nor hypotheses. Although, the Tablet View displays bookmarked states of visualizations that can be organized, annotated and connected manually. This can be used to structure findings, derive insights, and connect insights to hypotheses which can be added to the view and refined by the user during the analysis process. This must be done manually, because Jigsaw does not provide any automated support of this process. Figure 6c shows an example for hypotheses validation using the Tablet View.

Support for the *knowledge generation loop* is challenging, because knowledge generation is a process done entirely by the user, and involves concepts like trust or reasoning. In addition to these human factors, user's domain knowledge plays an important role, which is hard to incorporate, because it is difficult to externalize. Depending

(a)   (b)   (c)
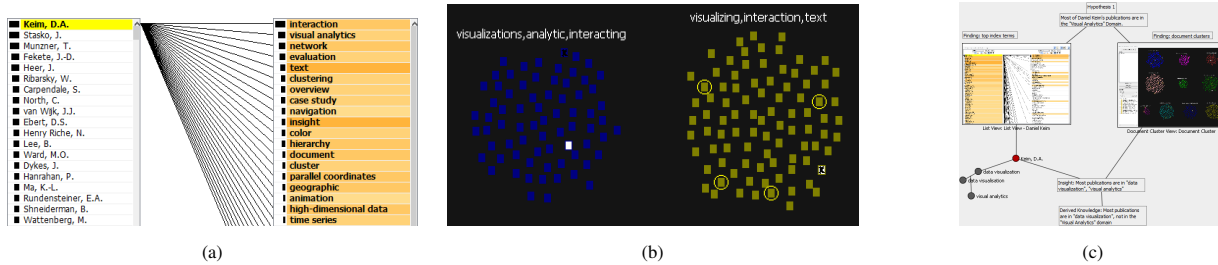
Fig. 6: Illustration of validation steps using Jigsaw. In 6a left, the person *Daniel Keim* (left list) is connected to the concept of *text*, displayed on the right. To find evidence for this fact, the Document Cluster View of the publications is opened (6b). After inspection of the cluster labels, a document from the *visualizing,interaction,text* cluster is examined in detail. This document is evidence that the fact presented in the list view in 6a is true. An example for the Tablet View can be seen in 6c.

on the nature of the problem, the required kind and necessary level of support varies. An example would be the automatic search and suitable presentation of further evidence for a given hypothesis, which has been already verified, but is not yet trusted enough to qualify as knowledge. If the Tablet View has been used during the verification loop to externalize the analysis process, the view can be helpful in the knowledge generation loop too.

We have showed that Jigsaw supports the three feedback loops which are part of the inductive reasoning process, and the amount of possible automated support varies.

Starting with the exploration loop, automated support is based mainly on predefined models, and is therefore limited by the analytic possibilities of the system. Going a step further to the verification loop, it gets harder to provide adequate automated support. The variety of different hypothesis types is an important reason for this. It is even more difficult to extend a system to support the knowledge generation loop, due to the increasing influence of human and other external factors, which cannot easily be learned and represented by computers. This makes it difficult to incorporate them in an automated process supporting knowledge generation. Further implications of loop automation are discussed in Section 5.3.

4.3   Comparative System Assessment

In this section, we provide a high level assessment of different systems from data mining, information visualization, visual analytics and provenance domain. The different natures of these systems illustrate the general applicability of our model on applications that deal with data, provide visualizations, and are designed to generate knowledge. The comparison is shown in Figure 7.

At first, we assess Knime [1](Version 2.9.1), which supports the interactive creation and execution of *data mining* pipelines. Data preparation and inspection, model building, model observation, and the model-vis mapping support is excellent, which is a good foundation for the exploration loops. The available data visualizations are on a basic level and allow brushing interactions only. Besides that, Knime provides no explicit support for the verification and knowledge generation loops, because there exist no tools to organize findings, derive insights, or connect hypothesis with insights. As a substitute, separate paths for the organization of findings and hypotheses can be added to an already existing pipeline. This is a possible way of recording the actions leading to a pipeline outcome, which can contribute to the knowledge generation process.

Next, we assess Tableau Desktop (Version 8.2 PE), which offers a number of visualizations that can be interactively adjusted by the user. Tableau is a representative of applications from the *information visualization* domain. Therefore, it has strong support for model-vis mapping, which is also the case for data preparation and data inspection. When it comes to model building, Tableau provides basic functionality. For specific data sets, it is possible to add a trend line or compute a forecast, where the model can be adjusted in designated bounds. The various manipulation and visualization options provide strong support for the exploration loops. Verification loops and knowledge generation loops support is also available with the story module, which allows free creation of reports and references to visualizations. In addition, all three loops are supported by the annotation tool, which allows the
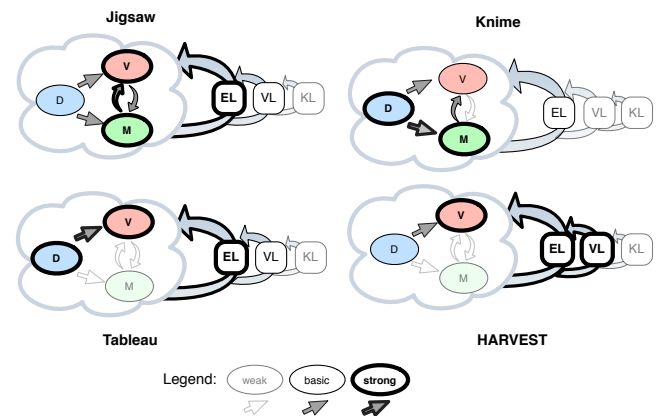

Fig. 7: Comparative model application to different kinds of systems. Jigsaw represents visual analytics, Knime data mining, Tableau information visualization, and HARVEST applications from the provenance domain. Strength of support for functionalities/components of our model is indicated by the weight of the lines.

addition of persistent annotations to data points or specific locations and areas in the visualization.

Jigsaw is an example for a *visual analytics* application, which we already examined in detail in Sections 4.1 and 4.2. It supports all the actions we included in our model. In addition, all three loops are supported, but there is some potential for improvements. For example, a good support of the verification loops can be easily achieved by a tighter integration of the Tablet View in the system. Also, providing histories of actions leading to a specific state of a visualization (bookmark) is a step towards better support of the verification and knowledge generation loops.

At last, we examine HARVEST [11, 37], a system which supports *provenance*. While using HARVEST, all interaction is recorded and can be used in order to understand the way findings have been detected. Using this data, the system is able to support analysts during the exploration and verification loops, for example by an automatic ranking of manually created notes based on the users behavior. This comes close to our definition of the knowledge generation process. The system also supports the analysis process by providing visualization recommendations or ordering of notes connected insights, or findings. Similar to Jigsaw's Tablet View, the note-taking-interface is capable of organizing, grouping, and ordering items, which supports the knowledge generation loops.

The assessment of the tools from different application domains shows that the model can be applied on applications, which work with data and visualization in general. The result also clearly separates the different application domains.

5   SUMMARY AND DISCUSSION

This study identifies new perspectives on visual analytic processes beyond weaving existing frameworks into one. Our model highlights that human and machine are a loop in the knowledge generation process using visual analytics. While existing models focus on one of the these,

our model integrates human thinking whilst describing visual analytics components. Most of the visual analytics systems do not fully cover or properly treat all aspects that our model requires as it includes the holistic process of knowledge generation which involves visual analytics systems and human users in the loop. Our model also specifies how human loops are intertwined with the subprocesses of visual analytics tools. We look at systems in a skeleton view using Keim's previous model and show how the interplay between each subcomponent can be influenced by human decision making and reasoning processes. This is important for researchers as it enables discussions of specific functions and their impacts on reasoning processes more explicitly because our model can describe the whole path from data to knowledge and vice versa. Besides connecting to system components, the model also defines human concepts and introduces three self-contained loops/stages of reasoning/thinking. We can also use our model to assess visual analytics system in terms of its functions toward human analytic outputs. For instance, we can detect areas of visual analytics systems that tend to cause biases within the knowledge generation process by aligning experts' analytic processes and outcomes against our model. Then, designers could improve corresponding visual analytics components to enhance early detection of such analytic failures: wrong hypotheses, conflicts between findings and insights, and dead ends of exploration cycles. We also find results that resonate with sensemaking, cognition and reasoning models (e.g. Pirolli and Card [29]). Our model even specifies where our current visual analytics systems fall short of, which is to support higher level loops, namely the exploration, verification and knowledge generation loops. Supporting higher level loops is more complex, however this would be a useful addition to many of the current systems [31].

5.1 Collaboration and Communication

Our model describes the knowledge generation processes for visual analytics. There are specific types of visual analytics, which our model does not explicitly mention. Visual analytics processes could be collaborative, so multiple stakeholders asynchronously or synchronously analyze data together and gain insights through verbal communication between them. Our model currently assumes an individual's analytic process, so it is missing collaborative components. We can simplify this collaboration process with different but possibly shared knowledge generation loops of all humans (i.e., white nodes in Figure 1). We can explain that a number of users perform actions and interpret findings together to improve the quality of their knowledge. Also distributed cognition theories have to be considered when examining representations and interactions among humans and artifacts (Liu et al. [20]). Nobarny et al. [24] already developed a system and performed studies focusing on distributed cognition in order to support collaborative visual analytics.

Especially the externalization and communication of information plays a crucial role that needs more detailed investigations with regards to the presented concepts of findings, insights or knowledge. The importance of externalization in visual analytic processes can be observed by heavy reliance on note-taking throughout analysis processes [22]. In real world, groups of users often take a variety of approaches to synthesize information with their own organizational language [32]. Communicating knowledge in visual analytics is actually the communication of evidence found in data supporting a belief. This evidence is shared with findings or insights, e.g., a commented visualization revealing an interesting relationship in data. The communication counterpart can follow the evidence and, depending on the level of trust, gain own insights or knowledge out of it. Collaborative analysis scenarios allow communication partners to verify this evidence with the system. We should also note that collaboration between multiple analysts open up chapters about maintaining exploration awareness [38]. In case of presentations or static documents the missing visual analytics system is replaced by the presenter or reporter. During presentations questions and answers are able to replace the knowledge generation loops. Static documents have to convey all information from hypotheses, findings, and insights in a way that readers can follow the conclusions. Interactive reports or documents (e.g., infographics) go beyond the

aforementioned scenario as they can be seen as a report combined with a limited visualization that is tailored to show the relevant findings of insights.

5.2 Visual Analytics of Streaming Data

In addition to collaborative visual analytics, we can apply our model to visual analytics of data streams (e.g., monitoring twitter data). In this situation, our data node is dynamically changing, so our model reformulates its visual representations and analytic models according to substantial changes made by the data streams. Further investigation is necessary since a number of requirements on data management, knowledge discovery and visualization need to be researched due to the dynamic nature of streams. Mansmann et al. [23] illustrate this and highlight that the analyst role changes since exploration is extended to include real-time monitoring tasks where situational awareness and complex decision making come into play.

5.3 Automatic Support of Knowledge Generation

We will now consider how novice users can be given guidance using the system and supporting analysis can aid knowledge generation. Systems can make automatic suggestions based on their current state, e.g., choosing a suitable color mapping or selecting parameters based on data.

5.3.1 Exploration Loop

The exploration loop is the basis of all knowledge generation in visual analytics. An important trigger to observe findings is how the system handles interactions. Analysts can learn from the causality between interactions and reaction, hence, supporting the exploration loop requires a system to respond with an immediate observable reaction to any interaction. Many algorithms in data analysis require complex computations and are not able to calculate a complete result immediately. In these cases, analysts should at least receive feedback that the algorithm is still running. Systems should provide the ability to switch between the states before and after calculation so analysts can learn from interactions. If the final result can be estimated, for instance, by intermediate results of an incremental algorithm, the estimation could be shown to the analysts with the additional ability to abort the calculation.

Findings are related to unexpected results or patterns in models or visualizations. Automatically detecting unexpected results is complicated, because it requires a definition of what analysts are expecting, which is usually not known to systems. On the other side, pattern mining algorithms are extracting patterns directly from data and in visualization, patterns could be detected with automatic methods as well. Even automatic methods to judge the usefulness of visualization exists, e.g., Bertini et al. [3] gives an overview of quality metrics approaches.

Actions are dependent on findings and the goal of the analysis. Visual analytic systems could provide suggestions for further actions in the analysis process, as in behavior-driven visualization recommendation [10]. Based on findings, these suggestions could offload some burden from analysts of having to choose the right action from all different possibilities. Novice users would benefit from such suggestions, because the system would present proper actions allowing users to learn the abilities of the system. For expert users the interaction with the system would be more efficient compared to navigating a large set of options. It is important to find an adequate level of suggestions based on user experience. A solution to this problem could be a learning technique, such as active learning (e.g., Settles [35]), that adapts the suggestions to users and minimizes the interaction costs.

5.3.2 Verification Loop

The verification loop is the central part in the knowledge generation loops. Analysts combine findings from data with their domain knowledge and gain new insights into the problem domain. The knowledge of analysts play an important role in the verification loop, therefore automatic support for the verification loop is limited to helping analysts record their analysis results.

Systems are not directly generating insights but analysts gain new insights from data when they are able to interpret findings. In this

respect systems can support this process by providing useful summaries of findings by allowing analysts to organize findings, hypotheses, or insights. Often insights are not dependent on a single finding but are hidden in complex relationships in the data and often difficult to find without prior knowledge. Systems addressing this problems are, for instance, Shrinivasan et al. [37] and Wright et al. [42].

To formulate hypotheses about the problem domain, findings are not enough as analysis requires insights into the domain. However, systems can automatically formulate hypotheses about the analyzed data from findings. Theses hypotheses can help analysts get a faster overview of unknown data sets but their use for complex analysis tasks is limited, because only hypotheses about relations in data can be generated automatically.

5.3.3 Knowledge Generation Loop

Automating the steps from insights to knowledge or from knowledge to new hypotheses is according to our definition not possible. Analysts gain new knowledge when the evidence collected with a visual analytics system is convincing. The best way to support knowledge generation with visual analytics are systems with the ability to look at data from different perspectives. This gives analysts the possibility to collect versatile evidence and increases the level of trust in findings or insights.

5.4 Future Investigations on Visual Analytics Systems

Our model specifies some areas that our current research can further investigate. We find some interaction types missing in many systems and interaction models, especially in the model construction or the coupling of models and visualizations. Many visual analytics tools tend to maintain preloaded models or to provide a very limited capability to manipulate models by adjusting some parameters. We observe that data become more and more dynamic and unstructured and human interaction on the model part is therefore crucial to analyze such data. Secondly, we see that many improvements could be made to further support human actions. Systems can actively learn from user behavior and adapt its models and visualizations, too. Visual analytics systems could proactively seek next candidate actions based upon user-generated logs. Our model points out that we need more explicit support to transfer findings, insights, and knowledge. As Endert et al. [7] states, visual analytics should recognize and integrate human working processes into the system. Different ways of how systems could offload some burdens from human users exist. The key is the interaction with visual analytics, although we should be aware of interaction costs (Lam [18], van Wijk [40]). Frequent interactions triggered by the system could demand too much effort, which may discourage user's exploration.

After analysis, the results have to be documented or communicated to others. An interesting ability of visual analytics systems could be a semi-automatic approach for generating documentations. Algorithms could detect interesting new findings and put them together in a convincing form, removing dead ends and duplicate findings. Such intelligent journals would make creation of reports or presentations much easier. Alternatively, written reports could be enriched with findings supporting statements in the document.

The analysis process often follows one direction and many findings are not explored in depth. Visual analytics systems could provide functionalities to backtrack former analysis results and suggest interesting but not investigated findings for further analysis. This would require algorithms to judge the interestingness of findings in context of the conducted analysis results. This functionality together with a short summary of previous results could be helpful for continuing an interrupted analysis session.

5.5 Real World Scenario

Our model reflects an ideal scenario where an analyst uses one single visual analytics system that is capable of handling all requirements but real world scenarios are different. On the one hand, users are often not aware of the systems capabilities and lack the required expertise to understand complex analysis methods (Gulf of Execution). On the other hand, the system's capabilities are not sufficient to solve an analysis task. As an outcome, analysts may stop their analysis in order to consult domain experts or continue their analysis with another system. In addition our model implies that each action ends in an observable reaction of the system. Real world systems often lack of this capability, which is a known problem in interaction science (Gulf of Evaluation).

The analysis of real world problems requires both expertise about the analysis and the domain. Domain experts often lack experience in understanding computer systems, visualization techniques and analysis methods, whereas visual analytics experts lack of sufficient domain knowledge. Thus, analysis requires collaboration between them. Without domain knowledge a visual analytics analysts is able to generate findings and insights concerning the data, not the domain. The domain expert is responsible for the formulation of problem hypotheses, the detection and interpretation of patterns. Domain experts need to be familiar with visual analytics methods and systems. On the other hand analysts have to learn about the problem domain.

One thing to keep in mind is that our model simplifies many different processes, and it contains inherent fuzziness, especially on the human side. That is the reason why our model consists of various loops that represent several levels of thinking. Obviously, no visual analytics tools can differentiate exactly between reasoning processes. Furthermore, the processes can disseminate many influences into other processes, which may not clearly appear in our model. Often many conflicting hypotheses are investigated in parallel and derived findings, insights or knowledge may affect each other.

5.6 Teaching Visual Analytics

Our model can be useful to provide a general overview for novice students, designers, and researchers in visual analytics. Teaching visual analytics often require teachers to provide fundamental concepts commonly appearing throughout applications, which inevitably involves domain knowledge. Our model can be used to explain the knowledge generation process without the need for such expert language. In addition, this model touches upon various components of visual analytics, such as interactions and automatic algorithms, in the perspectives of visual analytics applications. This guideline could point researchers to relevant literature in case they want to find out about specific methods. Our model also highlights the importance of the interplay and collaboration between human and machines. This rather obvious but easily forgotten notion could be highlighted, as illustrated in Figure 1.

6 Conclusion

In this paper we present a process model for knowledge generation in visual analytics that integrates system and human aspects. The model defines and relates relevant concepts and provides a knowledge generation process from knowledge to data and vice versa. The model embeds concepts into a three loop framework and illustrates possible human machine pairings that are fundamental in visual analytics. We illustrated that our model integrates with existing models and theories that focus on specific parts within the overall context. We demonstrated the models application using Jigsaw as an example system as well as undertaking a comparative assessment of three other well-known applications. Finally, we discussed model implications, named open issues, and pointed to future directions. The right hand side (human part) of our model is not restricted to visual analytics and can also be relevant for other disciplines as it combines computer and human based theories. The model aims to give a basis for more detailed compositions of theories. Whilst it is not our intention to cover every single details of such processes, we do provide a overview that commonly appears in visual analytics processes. We also acknowledge that human's knowledge processes cannot be linear or clearly subdivided into components of our model (e.g., insight). However, we provide inherently limited but meaningful distinction between human's knowledge gaining processes.

References

[1] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, and B. Wiswedel. Knime - the konstanz information miner: version 2.0 and beyond. *SIGKDD Explorations*, 11(1):26–31, 2009.

[2] E. Bertini and D. Lalanne. 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*, pages 12–20. ACM, 2009.

[3] E. Bertini, A. Tatu, and D. A. Keim. Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2203–2212, 2011.

[4] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2376–2385, 2013.

[5] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[6] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *Computer Graphics and Applications, IEEE*, 29(2):14–17, 2009.

[7] A. Endert, M. 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*, pages 1–25, 2014.

[8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37, 1996.

[9] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. T. Stasko. Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1646–1663, 2013.

[10] D. Gotz and Z. Wen. Behavior-driven visualization recommendation. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, page 315324, New York, NY, USA, 2009. ACM.

[11] D. Gotz, Z. When, J. Lu, P. Kissa, N. Cao, W. H. Qian, S. X. Liu, and M. X. Zhou. Harvest: An intelligent visual analytic tool for the masses. In *Proceedings of the First International Workshop on Intelligent Visual Interfaces for Text Analysis*, IVITA '10, pages 1–4, New York, NY, USA, 2010. ACM.

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

[13] T. Green, W. Ribarsky, and B. Fisher. Visual analytics for complex concepts using a human cognition model. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pages 91–98, Oct 2008.

[14] T. M. Green, W. Ribarsky, and B. Fisher. Building and applying a human cognition model for visual analytics. *Information Visualization*, 8(1):1–13, Jan. 2009.

[15] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann. *Mastering the Information Age - Solving Problems with Visual Analytics*. Eurographics Association, 2010.

[16] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. *Visual analytics: Scope and challenges*. Springer, 2008.

[17] B. C. Kwon, B. Fisher, and J. S. Yi. Visual analytic roadblocks for novice investigators. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 3–11, 2011.

[18] H. Lam. A framework of interaction costs in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1149–1156, Nov 2008.

[19] P. Legrenzi, V. Girotto, and P. N. Johnson-Laird. Focussing in reasoning and decision making. *Cognition*, 49(1):37–66, 1993.

[20] Z. Liu, N. J. Nersessian, and J. T. Stasko. Distributed cognition as a theoretical framework for information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1173–1180, 2008.

[21] L. Magnani. *Abduction, Reason and Science: Processes of Discovery and Explanation*. Springer, 2001.

[22] N. Mahyar, A. Sarvghad, and M. Tory. A closer look at note taking in the co-located collaborative visual analytics process. In *2010 IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 171–178, Oct. 2010.

[23] F. Mansmann, F. Fischer, and D. A. Keim. Dynamic visual analytics facing the real-time challenge. In *Expanding the Frontiers of Visual Analytics and Visualization*, pages 69–80. Springer, 2012.

[24] S. Nobarany, M. Haraty, and B. Fisher. Facilitating the reuse process in distributed collaboration: a distributed cognition approach. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1223–1232. ACM, 2012.

[25] D. Norman. *The Design of Everyday Things*. Basic Books, 2002.

[26] C. North. Toward measuring visualization insight. *Computer Graphics and Applications, IEEE*, 26(3):6–9, 2006.

[27] C. S. Peirce. *Collected papers: Elements of Logic*, volume 2, pages 56–59. Belknap Press, 1965.

[28] W. A. Pike, J. Stasko, R. Chang, and T. A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009.

[29] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, volume 5, pages 2–4, 2005.

[30] M. Pohl, M. Smuc, and E. Mayr. The user puzzle: Explaining the interaction with visual analytics systems. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2908–2916, Dec 2012.

[31] J. C. Roberts, D. A. Keim, T. Hanratty, R. R. Rowlingson, R. Walker, M. Hall, Z. Jacobson, V. Lavigne, C. Rooney, and M. Varga. From Ill-Defined Problems to Informed Decisions. *EuroVis Workshop on Visual Analytics (2014)*, 2014.

[32] A. Robinson. Collaborative synthesis of visual analytic results. In *IEEE Symposium on Visual Analytics Science and Technology, 2008. VAST '08*, pages 67–74, 2008.

[33] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 11(4):443–456, 2005.

[34] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2366–2375, 2013.

[35] B. Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52:55–66, 2010.

[36] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[37] Y. Shrinivasan, D. Gotz, and J. Lu. Connecting the dots in visual analysis. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 123–130, Oct 2009.

[38] Y. Shrinivasan and J. van Wijk. Supporting exploration awareness in information visualization. *IEEE Computer Graphics and Applications*, 29(5):34–43, Sept. 2009.

[39] J. T. Stasko, C. Görg, Z. Liu, and K. Singhal. Jigsaw: Supporting Investigative Analysis through Interactive Visualization. In *IEEE VAST*, pages 131–138. IEEE, 2007.

[40] J. van Wijk. The value of visualization. In *Visualization, 2005. VIS 05. IEEE*, pages 79–86, Oct 2005.

[41] T. von Landesberger, S. Fiebig, S. Bremm, A. Kuijper, and D. W. Fellner. Interaction taxonomy for tracking of user actions in visual analytics applications. In *Handbook of Human Centric Visualization*, pages 653–670. Springer, 2014.

[42] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis: Concepts and methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 801–810, New York, NY, USA, 2006. ACM.

[43] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, page 4. ACM, 2008.