Externalization of Data Analytics Models:

Toward Human-Centered Visual Analytics

Arman Didandeh[1(✉)] and Kamran Sedig[1,2]

[1] Department of Computer Science, Western University, London, ON, Canada
{adidande, kamrans}@uwo.ca
[2] Faculty of Information and Media Studies, Western University, London, ON, Canada

Abstract. Visual analytics tools (VATs) can support the execution of complex cognitive activities. Most VATs make use of analytics models in the execution of data-intensive activities. However, due to their non-transparent and black box nature, analytics models are hard to use. This leads to a lower degree of understandability of the models, which then results in a lack of trust and applicability. To overcome this problem and to help bring VATs closer in line with users' mental models, we introduce a framework for the externalization of analytics models in VATs. In doing so, we propose the use of scaffolding to provide the users with visual support in the execution of their cognitive activities. The aim is to make VATs more human-centered. To demonstrate the application of the externalization framework, we present two components of a visual analytics tool named VARSITY.

Keywords: Visual analytics · Cognitive activity · Scaffolding · Externalization · Human-centered computing

1 Introduction

Human-centered computing focuses on research, design, implementation, and evaluation of tools that best suit the perceptual and cognitive needs and abilities of humans to support the tasks and activities that they perform (Huang 2014; Kerren et al. 2006). This paper is concerned with human-centered visual analytics tools (VATs) that support the execution of data-intensive complex cognitive activities (Knauf and Wolf 2010; Sedig and Parsons 2013), such as analyzing financial markets. As the volume, velocity, types, and forms of data multiply, the complexity of data spaces with which users work grows exponentially, making the execution of the complex activities even harder. To overcome these issues, VATs combine automated data analytics and interactive visualization techniques to leverage both the reasoning abilities of humans and the powerful data discovery and modeling strengths of computers (Keim et al. 2008). Accordingly, VATs support distributed cognition (Pohl et al. 2012) by creating a joint cognitive system (Parsons et al. 2015), which reduces the cognitive load of users (Keim et al. 2008). VATs make use of analytics models (e.g., regression analysis, association rule analysis, dimensionality reduction) to enable users to reify the data through engaging with visualizations, making them the epistemic loci of VATs

S. Yamamoto (Ed.): HIMI 2016, Part I, LNCS 9734, pp. 103–114, 2016.
DOI: 10.1007/978-3-319-40349-6_11

(Sedig et al. 2012). In doing so, VATs should provide affordances so that the user can engage with the analytics models through the visualizations (Endert et al. 2012; Sacha et al. 2014).

Interactive steering of analytics models through the given visualizations is not trivial, as the users often view these models as non-transparent black boxes whose behaviors are not easy to understand (Cortez and Embrechts 2013; Keim et al. 2015). If a model is already built without the users being engaged in the process, then interpreting "*what the model is doing*" or understanding the model's application is not easy, resulting in a lack of trust in how the model's outcomes are obtained, in addition to its credibility and applicability. This lack of understandability has a huge impact on final results and the users' hypotheses and conclusions, affecting overall activity performance negatively—e.g., a bioinformatician, using the default parameters of an SVM model to classify clinical outcomes based on their gene signatures, experiences over-fitting in the form of too many false negatives, and thus decides to move on to another less effective classification model. Hence, to make analytics models human-centered, they must be brought closer in line with users' mental models (Endert et al. 2012). Two broad strategies have been suggested in literature: (i) visualizing analytics models to make them more transparent (Bradel et al. 2014; Holzinger and Pasi 2013); and (ii) making analytics models semi-automatic by allowing users to interactively select and adjust their data processing parameters, steer them, and construct their functionality (Keim et al. 2015; Ltifi et al. 2013). However, designing these such that they fit the perceptual and cognitive tasks of the users and create a coupling between analytics models and visualizations is challenging (Sacha et al. 2014) because the design must consider the visual perception and analytical abilities of the users (Kohlhammer et al. 2011). As well, to prevent an increase in the users' mental load when interacting with VATs (Keim et al. 2008), the design must aim to facilitate the cognitive and perceptual processes of the users (Knauff and Wolf 2010; Sedig and Parsons 2013). Making analytics models transparent and conceptually accessible to users requires research and the development of new visualization and interaction techniques that couple them to users' cognitive needs and tasks.

In this paper, to overcome the aforementioned impediments in complex cognitive activities, we briefly introduce the notion of externalizing analytics models through visualization techniques. This approach leads to a more human-centered analytics process, and hence a more human-centered design. To do so, there is a need for collaboration with researchers of disciplines other than visual analytics (Keim et al. 2012, 2015). As a result, users can explicitly engage with the analytics models—not solely with the visualizations of the raw data. To provide a framework for how to externalize analytics models, we briefly introduce an abstract formalization of the models. We also focus on the importance of a human-centered design. In doing so, we employ *scaffolding*, i.e., a human-centered design approach in visualization that provides the users with explicit access to both the data and the analytics spaces, allowing them to construct and evaluate data-based hypotheses. We recommend the proper use of interactive visualizations to gain insight into the analytics model, along with interaction as a means for communicating outputs of the models, for the purposes of education, explanation, and evaluation. Through a scaffolding process, the burdens of engaging with the analytics models, both in the form of model comprehension and model construction can be minimized.

As a testbed for showcasing the externalization framework, we present a brief description of a VAT named VARSITY (*Visual **A**nalytics of university **R**esearch network**S** and **I**ndus**T**r**Y** collaborations*). We provide two examples of VARSITY's analytics components which are operationalized according to the externalization framework. These components are: (1) a frequent itemset mining model, and (2) a correlation investigation model (which partially incorporates a textual topic extraction model). These two components are examples of our users' manifold interests in performing data-intensive activities using VARSITY.

The remainder of this paper is structured as follows. The next section provides some terminological and conceptual background to the reader. Then in Sect. 3, we focus on the idea of externalizing analytics models. Next, we briefly introduce the two components of VARSITY to showcase the externalization framework. Finally, Sect. 5 includes a summary and some future ideas.

2 Background and Terminology

2.1 Analytics and Analytics Models

The use of analytics has become popular in different areas, such as business (Chen et al. 2012) and security (Mahmood and Afzal 2013). However, there is a lack of a generally agreed-upon definition for the term. From an etymological standpoint, the suffix *-ics* denotes "the science and/or art of studying something[1]". Therefore, broadly, analytics can be defined as the science and/or art of studying the principles of analysis—and particularly that of the various forms of data within an ecosystem—to be used as an asset in activities such as investigation of underlying patterns and discovery of anomalous behavior. With this definition, an *analytics model* is a computational data model that is built upon a subset of the data space known as a dataset. Techniques and concepts that are incorporated to build such models, specifically in visual analytics, are usually imported from areas such as machine learning, data mining, knowledge translation, and statistics. Since analytics models reduce users' cognitive load and save them time and effort, their use in VATs is strongly advocated. Analytics models are often constructed and utilized in an iterative and multi-step process of the progression of a data-intensive cognitive activity. This might lead to a discourse—i.e., a back and forth communication process between the user and the tool (Keim et al. 2008). This process needs to be transparent (Kohlhammer et al. 2011) to be understandable. On this note, analytics discourse is a process involving the user, the user's tasks, the tool, and the data (in both the raw and intermediary forms and as output). The cognitive activity emerges from the whole process (Sedig and Parsons 2013).

As analytics models support different tasks, they vary in their functionality—e.g., prediction models vs. statistical analysis models. Some models may require intermediary steps in which the initial state of the raw data—either slightly or considerably—changes,

[1] http://www.etymonline.com.

resulting in *intermediary data*. Models also vary in complexity—e.g. from a simple model of a sorting algorithm to a fairly complex Bayesian dimensionality reduction model. As their complexity increases, and due to their high dependence on complicated mathematical concepts, they have been dealt with using black box representations (Cortez and Embrechts 2013; Keim et al. 2015). The tools that feature these models usually ask the user for the data upon which the model is to be built, along with some initial setup. Instead of a stepwise construction process, these tools then jump to the fully constructed model. Thus, it is only after the model is built that the user can engage with it—a highly non-optimal and non-trivial action (Keim et al. 2008) with negative effect on the overall execution of the cognitive activity (Wong et al. 2012). Furthermore, many of the state-of-the-art visual analytic techniques and algorithms create results that are unclear to the users, or do not incorporate their preferences, making these techniques less adaptable (Wong et al. 2012). In addition, for models that deal with huge amounts of data, the construction time of the models is a cost to the system. Thus, the understandability of the models might become a huge burden to their users[2]. In particular, this can happen when analytics models are being applied in domains different from mathematics and computer science, such as healthcare, insurance, and security (Meyfroidt et al. 2009). As a result, as the engagement of the users with the models diminish, the level of trust that they have in the outcome of those models dwindles.

Having analytics models paired with visualization techniques to support activities such as learning or decision-making is not a novel idea (Kerren et al. 2006). However, researchers have recently suggested that the design of analytics models be reconsidered (Keim et al. 2012, 2015). As an instance, algorithms must be re-designed, so that they can learn from their users (Endert et al. 2014). Moreover, since the details of analytics models are not well-known among professionals of other communities, there needs to be a way of communicating them (not only their output) without too much complexity (Meyfroidt et al. 2009). We propose that as part of this re-design process, analytics models and interactive visualizations can be paired, which in turn can reduce the disconnectivity between them. In order to do so, we propose the use of *scaffolded design* in externalizing the analytics models within a VAT.

2.2 Scaffolding and Scaffolded Design

Scaffolding is a process that supports the communication of knowledge through guiding, configuring, and/or disciplining a human activity (Alexander et al. 2013). In learning, it is described as “the support given during a learning process” so that learners engage in deeper aspects of information (Marai 2015). In visualization tools, scaffolding is a process in which visual support for a concept gradually transitions “from an intuitive stage of understanding to a reflective one”, promoting a higher degree of thought (Sedig et al. 2001). By providing deeper accessibility to embedded concepts and information, scaffolding affects attentive processes, mental load, learning, and decision-making activities of users positively (Alexander et al. 2013). Scaffolded

[2] There is an exception for the non-expert users who simply want to use the output of a model. These users are not the focus of this paper.

design is a highly human-centered approach (Quintana et al. 2004), as it minimizes the efforts of the users in interpreting the concepts, and enables users to act independently after mastering the tasks (Jin and Kim 2015).

While scaffolds are visual constructs that provide support and facilitate a user's tasks, scaffolding is the process of providing this support. During the scaffolding process, scaffolds might change, either slightly or drastically, gradually or abruptly, or even remain the same, depending on the user and the activity. Scaffolding is concerned with the design and operationalization of such scaffolds. For instance in analytics models, the characteristics, permanency, and existence of scaffolds that support them might: (1) remain at almost the same level, but change in time—e.g., clusters get updated during a clustering process, but the number of cluster scaffolds might stay the same; (2) increase in quantity—e.g., the cardinality of the set of all frequent itemsets increases throughout the mining process; (3) increase in quality—e.g., association rules are refined throughout their construction process; or (4) decrease in quantity and/or quality—e.g., reduction of visualization structures in progressive elaboration of concepts (Sedig et al. 2001).

Proper design of VATs can result in *human-centered fusion*, human cognition can be engaged, as it both provides support to and receives support from computational processing units (Hall and Jordan 2014). Therefore, it is very important which scaffolding processes are selected for the analytics models. Proper scaffolding can incorporate visual thinking, improve the quality of communication, understandability, and trust. Conjointly, scaffolds may be used as a means of providing feedback. This feedback can be from the tool towards the user through visual perception, or from the user towards the tool, through interaction with visualizations. Hence, scaffolds are good candidates when it comes to externalizing analytics models. In the next section, we focus on incorporating scaffolding in externalization of analytics models.

3 Externalization of Analytics Models

Apart from their outcome, not enough attention is paid to the details of analytics models. This is because many visual analytics experts have mainly focused on the presentation of visual results (Kohlhammer et al. 2011; Marai 2015). As the field advances, end-users are to have a higher degree of engagement with VATs. Thus, designers and researchers have advocated for data-related frameworks, tools, and techniques to adjust and be more human-centered (Endert et al. 2012). In the case of analytics processes, three levels for the involvement of users with the model have been proposed: *automatic* (no control), *user-driven* (partial control), and *user-steered* (full control) analytics (Von Landesberger et al. 2011).

The full control in user-steered analytics is non-trivial and often requires a great deal of effort from the user. Therefore, there is a need for the visual explication of analytics models—i.e., bringing them to the foreground of the analytics process and making them more usable Zhou and Chen (2015). However, when it comes to the construction and improvement of these models, users need to have a certain level of knowledge about their structure and functionality. Since a majority of the users of these models lack such expertise, they might find it hard to engage with the models using current interfaces. This still leaves the users mostly as consumers of the models, resulting in a harder interpretation of the models' application. This results in a lack of understandability of, and consequently a lack of trust, in the results of the analytics models, particularly in the case of complicated models.

We propose the externalization of the analytics models using visual scaffolding. The goal is to help users overcome the aforementioned problems with the understanding of and trust in analytics models and hence support them in the execution of complex cognitive activities. Scaffolding is intended to result in a more balanced information processing load between a VAT and its users as well as smoother flow of interdependent cognitive tasks. Additionally, users with different levels of expertise are meant to have better engagement with the tool, resulting in a more transparent analytics process (Kohlhammer et al. 2011).

In order to externalize analytics models, models of different functionalities and complexity levels need to be described. We propose a simple framework for describing and representing analytics models. The framework conceptualizes models in terms of 5 generic and abstract features: (1) *input*—they receive data; (2) *behavior*—they manifest behavior according to their underlying algorithms; (3) *parameters*—their behavior can be manipulated by adjusting certain internal parameters; (4) *steps*—they generate intermediary data; and (5) *output*—they produce output data. To increase the degree of engagement of users with analytics models, we can use these features and map them onto visual scaffolds. For this mapping, we use two scaffolding techniques: *interpretive* and *constructional*. Interpretive scaffolding mostly uses the externalization of an analytics model to support users with comprehension and understanding. This type of scaffolding helps users engage with scaffolds that represent a model's behavior. Constructional scaffolding requires a higher degree of user engagement. It needs participation in model construction. Both techniques are intended to enable users to engage with externalized representations of the 5 features of analytics models. Each of these features needs to be properly scaffolded according to interactivity considerations (Sedig et al. 2012). To demonstrate an application of our feature-based framework, we briefly present two components of a tool, VARSITY, in the following section.

4 VARSITY

VARSITY is a VAT designed for university administrators (e.g., deans, department chairs, and other stakeholders) to help them analyze research networks and industry collaboration data. Among the many tasks and activities it supports, VARSITY is aimed at helping with decision-making activities involving awards and research directions. Specifically, it helps with gaining insight into the existing large body of data, learning about hidden patterns in data, and, ultimately, making strategic administrative decisions.

Due to space constraints, we do not provide a detailed description of VARSITY's design and implementation. Here, we only provide a brief overview of the technical issues for interested readers. VARSITY runs on a Node.js web server and the implementation follows an AngularJS model-view-controller (MVC) framework. The data space includes publication, awards (i.e., grants), and faculty members' data.

Publications data are retrieved from the Elsevier Scopus API, and the latter two are provided by the university. The data are processed using scripts and are subsequently stored in a MySQL database. The client modules ask for data through secure requests and receive them in JSON format. Our client side is implemented using HTML5, CSS3, and jQuery. We use D3.js to encode data to SVG elements.

Although the interests of the users of VARSITY are manifold, in this paper we focus on two specific activities: (1) detection and investigation of the co-occurring topics among faculty members' publications, and (2) exploration of correlations between publications and awards, and examination of the potential causal relationships among them. These activities and their corresponding implementations are described below.

4.1 Frequent Topic-Sets

The publications data enables administrators to investigate the university's overall research agenda as well as detect the emerging research directions. Publications can be classified in a multi-class setting[3] using 334 distinct topics and 27 topic groups, which are standardized by Elsevier. The topic groups can resemble research disciplines to an accurate degree—e.g. Arts, Business, Computer Science, and Immunology. To detect co-occurring topics, we modified the *frequent itemset mining* model, mainly used for market-basket analysis (Moens et al. 2013). Each publication can be represented as a container (AKA: basket) that includes a number of topics (AKA: items). A topic-set is a group of topics that a number of publications share. Thus, each topic-set corresponds to a numeric value known as the support. Frequent topic-sets are those with a support higher than a threshold. To find the frequent topic-sets, we mine the space of publications with at least one author being a university affiliate.

The input data space can be filtered according to publication attributes such as publication date and the home faculty/department of the authors. Three different behaviors are supported in the current version of VARSITY; the user can choose to see: (1) the overall frequent topic-sets, (2) a specific topic group (i.e., discipline) and mine for frequent topic-sets within, or (3) a number of topic groups to have a directed, yet integrated cross-discipline investigation. In terms of the intermediary data, each topic-set corresponds to a subset of the publication space. These publications can be further investigated as a unit, or individually. As for the parameters (e.g., choice of noise removal method or data structures being used), this analytics model uses a threshold to filter the topic-sets to return the frequent ones.[4] Due to the modified behavior of our model, we no longer incorporate an explicit threshold value. This is as a consequence of the tasks of the user, as in certain scenarios even non-frequent topic-sets also contain valuable information. For example, one might be interested in the fact that recently *Computer Science* and *Immunology* have not co-occurred frequently, and *Infectious Diseases* has replaced *Immunology* in the last five years. Last,

---

[3] In a multi-class setting, each item can be classified as more than one class, as classes are parallel rather than complementary.

[4] Here, we are concerned with internal implementation-level parameters.

Fig. 1. Frequent topic-sets for *Pharmacology, Toxicology, and Pharmaceutics* under interpretive scaffolding mode: some topics do not co-occur in groups of 4.

but not least, is the output of the model. For the sake of demonstration in this paper, and as a proof of concept, we have limited the maximum number of topics within a topic-set to five. This is because empty topic-sets are also of interest to users, and the total number of possible topic-sets grows factorially.

For this component of VARSITY, we have considered both the interpretive as well as the constructional scaffolding techniques. In the former case, users are provided with support in terms of understanding and—a certain degree of—engagement with the frequent topic-set model. In the case of the latter, users have access to visual affordances that enable them to manually construct different topic-sets according to their needs, hypotheses, and thought processes. These affordances are in the form of empty placeholders for topics and topic-sets and create a dynamic and back-and-forth dialogue between the user and the analytics model. Figure 1 demonstrates the functionality of the component in the interpretive scaffolding mode. The opacity of the topic-set circles encodes their support within their level, while the presence or absence of topics (encoded in different shapes) is demonstrated using color-coding.

4.2 Correlation-to-Causality

Finding correlations is a popular task in visual analytics (Pfaffelmoser and Westermann 2013; Kay and Heer 2016). It usually includes analyzing data items of the same nature—i.e., different instances of the same entity within a dataset. However, when it comes to finding correlations between instances of different entities, and to the best of our knowledge, there has not been a great deal of research in the visual analytics community. For this component of VARSITY, our users are interested in finding potential correlations between publications and awards. This enables them to hypothesize and investigate potential cause-and-effect between the two—i.e., whether awards have led to publications or vice versa. To do so, we have implemented an analytics model that mines the joint space of publications and awards. The model links publications and awards based on attributes such as shared authors/award investigators, their temporal attributes, topics, keywords, and so on.

Our analytics model provides users with three levels of relaxation in the correlation mining process. This is intended to help the model support different degrees of generalization. In the first mode (very relaxed), publications with authors who are also award investigators (whether primary or co-investigators) are detected (the author correlation phase). This is non-trivial due to different formats and language settings of names for the different sources from which the initial data comes. Thus, we have used computational linguistics algorithms, which due to space constraints are not described here. Next, a keyword correlation investigation takes place, using a keyword matching algorithm. A confidence level and a choice of correlation score calculation (either uniform or weighted) are set. The very relaxed mode of the analytics model returns all the possible publications that have correlations with a selected award, regardless of a lack of keyword correlation. The second level of relaxation, however, only returns the publications that have keyword correlations. Therefore, the analysis is less relaxed and focuses on relationships with more possibility of causal detection. To add a new layer of correlation between the awards and the publications, the third level of relaxation (specific mode) incorporates a topic extraction algorithm based on the LDA technique. Topics are represented as a cluster of co-occurring terms, which demonstrate the underlying themes of the documents (Chuang et al. 2012).

In terms of visual support, VARSITY provides users with a treemap of awards (encoded with size according to their amounts) within various departments (encoded with colors). The ability to filter the awards based on different attributes helps users choose the award to be investigated. Upon selection of an award, the user is provided with visual controls to steer the analytics model and analyze both the intermediary data and the output in a back-and-forth manner throughout the analytics discourse. Clues to possible next steps in the analytics discourse are initially visually scaffolded and then disappear gradually; however, they remain available to the user on demand. The correlated publications are grouped based on their years, and a correlation score is provided for each, using a normalized bar. Moreover, the presence or absence of specific keywords/topic terms is scaffolded as well. Furthermore, one can investigate publications in terms of their standard topic groups and topics, as well as their authors. University authors are distinctly visualized to suggest further investigation. For publications with too many authors, further investigation is separated and scaffolded onto a new layer on top of the main canvas. Figure 2 shows the results for the relaxed level of this analytics model. Correlated publications are represented as circles and their correlation score as a bar to their left. Award keywords are color-coded and their existence within a publication is represented as a pie inside the publication circle. Authors of the publications are also shown on demand, along with the topics of the selectd publications. The user can steer the model and analyze both the intermediary and output data throughout the analytics discourse.

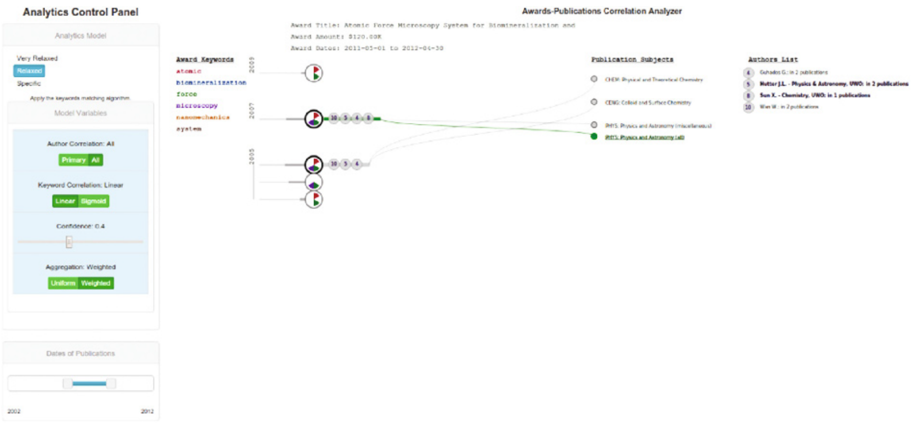

Fig. 2. The output and partial intermediary data of the relaxed mode of the Correlation-to-Causality model for a selected award. Some affordances are activated on demand.

5 Summary and Future Work

The use of analytics models is strongly advocated within the visual analytics community. Yet, due to their complicated nature and standard non-transparent behavior, their understandability, adaptability, applicability, and trust in their outcome have been fairly restricted, specifically when given to non-expert users with little knowledge of the underlying structures and behaviors of these models. In addition, there has been a number of concerns about the lack of a human-centered approach in the design of these tools when it comes to their analytics components. In this paper we introduced a framework for the externalization of analytics models through visualizations. Using human-centered design principles, and in particular, visual scaffolding, designers can externalize different analytics models within VATs. This enables the users of the tools to have a higher degree of engagement with the models through interactive externalized analytics models. Hence, users will be more involved with the analytics discourse, and thus the level of cognitive load in the execution of complex cognitive activities will decrease. Our proposed framework is comprised of 5 general features of the analytics models. These features can be externalized using scaffolding techniques to provide visual support to the users. We provided two examples to demonstrate the feasibility of this framework and the techniques. In future work, we plan to add more elaborate analytics models to VARSITY and solicit feedback from end-users to improve the framework. We also plan on investigating interactivity concerns of the externalization framework.

References

Alexander, E., Eppler, M.J., Bresciani, S.: Knowledge scaffolding: a classification of visual structures for knowledge communication in teams. In: Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies, p. 8. ACM, September 2013

Bradel, L, North, C, House, L., Leman, S: Multi-model semantic interaction for text analytics. In: Proceedings of IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 163–172 (2014)

Cortez, P., Embrechts, M.J.: Using sensitivity analysis and visualization techniques to open black box data mining models. Inf. Sci. **225**, 1–17 (2013)

Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**(4), 1165–1188 (2012)

Chuang, J., Manning, C.D., Heer, J.: Termite: visualization techniques for assessing textual topic models. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 74–77. ACM, May 2012

Endert, A., Fiaux, P., North, C.: Semantic interaction for sensemaking: inferring analytical reasoning for model steering. IEEE Trans. Vis. Comp. Graph. **18**(12), 2879–2888 (2012)

Endert, A., Hossain, M.S., Ramakrishnan, N., North, C., Fiaux, P., Andrews, C.: The human is the loop: new directions for visual analytics. J. Intell. Inf. Syst. **43**(3), 411–435 (2014)

Hall, D.L., Jordan, J.M.: Human-Centered Information Fusion. Artech House, Norwood (2014)

Holzinger, A.: Human-computer interaction and knowledge discovery (HCI-KDD): what is the benefit of bringing those two fields to work together? In: Cuzzocrea, A., Kittl, C., Simos, D. E., Weippl, E., Xu L. (eds.) CD-ARES 2013.LNCS, vol. 8127, pp. 319-328. Springer, Heidelberg (2013)

Huang, T.W. (ed.): Handbook of Human Centric Visualization: Theories, Methodologies, and Case Studies. Springer, Heidelberg (2014)

Jin, S.H., Kim, T.H.: Visual scaffolding for encouraging online participation. In: 2015 International Conference on Interactive Collaborative Learning (ICL), pp. 685–688. IEEE, September 2015

Kay, M., Heer, J.: Beyond Weber's law: a second look at ranking visualizations of correlation. IEEE Trans. Vis. Comput. Graph. **22**(1), 469–478 (2016)

Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Human-Centered Issues and Perspectives. Lecture Notes in Computer Science, vol. 4950, pp. 154–175. Springer, Heidelberg (2008)

Keim, D.A., Munzner, T., Rossi, F., Verleysen, M.: Bridging information visualization with machine learning (Dagstuhl Seminar 15101). Dagstuhl Reports, vol. 5, no. 3 (2015)

Keim, D.A., Rossi, F., Seidl, T., Verleysen, M., Wrobel, S.: Dagstuhl manifesto: information visualization, visual data mining and machine learning. Informatik-Spektrum **35**(4), 311–317 (2012)

Kerren, A., Ebert, A., Meyer, J.: Human-Centered Visualization Environments. Springer, Heidelberg (2006)

Knauff, M., Wolf, A.G.: Complex cognition: the science of human reasoning, problem-solving, and decision-making. Cogn. Process. **11**(2), 99–102 (2010)

Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., Andrienko, G.: Solving problems with visual analytics. Proc. Comput. Sci. **7**, 117–120 (2011)

Ltifi, H., Kolski, C., Ayed, M.B., Alimi, A.M.: A human-centred design approach for developing dynamic decision support systems. J. Decis. Syst. **22**(2), 69–96 (2013)

Mahmood, T., Afzal, U.: Security analytics: big data analytics for cybersecurity: a review of trends, techniques and tools. In: 2013 2nd National Conference on Information Assurance (Ncia), pp. 129–134. IEEE, December 2013

Marai, G.E.: Visual scaffolding in integrated spatial and nonspatial analysis. In: EuroVis Workshop on Visual Analytics (EuroVA). The Eurographics Association (2015)

Meyfroidt, G., Güiza, F., Ramon, J., Bruynooghe, M.: Machine learning techniques to examine large patient databases. Best Pract. Res. Clin. Anaesthesiol. **23**(1), 127–143 (2009)

Moens, S., Aksehirli, E., Goethals, B.: Frequent itemset mining for big data. In: 2013 IEEE International Conference on Big Data, pp. 111–118. IEEE, October 2013

Pfaffelmoser, T., Westermann, R.: Correlation visualization for structural uncertainty analysis. Int. J. Uncertain. Quantif. **3**(2), 171–186 (2013)

Parsons, P., Sedig, K., Didandeh, A., Khosravi, A.: Interactivity in visual analytics: use of conceptual frameworks to support human-centered design of a decision-support tool. In: 2015 48th Hawaii International Conference on System Sciences (HICSS), pp. 1138–1147. IEEE, January 2015

Pohl, M., Smuc, M., Mayr, E.: The user puzzle—explaining the interaction with visual analytics systems. IEEE Trans. Vis. Comput. Graph. **18**(12), 2908–2916 (2012)

Quintana, C., Reiser, B.J., Davis, E.A., Krajcik, J., Fretz, E., Duncan, R.G., Kyza, E., Edelson, D., Soloway, E.: A scaffolding design framework for software to support science inquiry. J. Learn. Sci. **13**(3), 337–386 (2004)

Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G., Keim, D.A.: Knowledge generation model for visual analytics. IEEE Trans. Vis. Comp. Graph. **20**(12), 1604–1613 (2014)

Sedig, K., Klawe, M., Westrom, M.: Role of interface manipulation style and scaffolding on cognition and concept learning in learnware. ACM Trans. Comput.-Hum. Interact. (TOCHI) **8**(1), 34–59 (2001)

Sedig, K., Parsons, P.: Interaction design for complex cognitive activities with visual representations: a pattern-based approach. AIS Trans. Hum.-Comput. Interact. **5**(2), 84–133 (2013)

Sedig, K., Parsons, P., Dittmer, M., Ola, O.: Beyond information access: support for complex cognitive activities in public health informatics tools. Online J. Public Health Inform. **4**(3) (2012)

Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J.J., Fekete, J.D., Fellner, D.W.: Visual analysis of large graphs: state-of-the-art and future research challenges. Comput. Graph. Forum **30**(6), 1719–1749 (2011). Blackwell Publishing Ltd.

Wong, P.C., Shen, H.W., Johnson, C.R., Chen, C., Ross, R.B.: The top 10 challenges in extreme-scale visual analytics. IEEE Comput. Graph. Appl. **32**(4), 63 (2012)

Zhou, J., Chen, F.: Making machine learning useable. Int. J. Intell. Syst. Technol. Appl. **14**(2), 91–109 (2015)