Human Factors in Information Visualization and Decision Support Systems

André Calero Valdez[1], Philipp Brauner[1], Martina Ziefle[1], Torsten Wolfgang Kuhlen[2], Michael Sedlmair[3]

Human-Computer Interaction Center, RWTH Aachen University[1]
Visual Computing Institute, RWTH Aachen University[2]
Visualization and Data Analysis Group, TU Wien[2]

Abstract

With the increase in data availability and data volume it becomes increasingly important to extract information and actionable knowledge from data. Information Visualization helps the user to understand data by utilizing vision as a relatively parallel input channel to the user's mind. Decision Support systems on the other hand help users in making information actionable, by suggesting beneficial decisions and presenting them in context. Both fields share a common need for understanding the interface between the computer and the human. This makes human factors research critical for both fields. Understanding limitations of human perception, cognition and action, as well as their variance must be understood to fully leverage information visualization and decision support. This article reflects on research agendas for investigating human factors in the aforementioned fields.

1 Too much data, too little time

Data is everywhere. The amount of collected and generated data every year exceeds the amount of all years before (McAfee et al., 2012). Through the pervasiveness of information and communication technology, data is created in almost all commercial and non-commercial processes, whether known by the user or not. Every step leaves a digital footprint, ready to be processed. But not only the amount of data is unprecedented. According to McAfee et al. (2012) three V's characterize what Big Data is. *Volume* as we have already is the most obvious driver of a need for new methods to handle Big Data. *Velocity* refers to the speed of how fast data is generated and how fast it needs to be processed. Real-time stock data needs evaluation now. Lastly, *variety* implies that data comes in various (un-)structured formats, but is at the same time heavily interconnected.

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
B. Weyers, A. Dittmar (Hrsg.):
Mensch und Computer 2016 – Workshopbeiträge, 4. - 7. September 2016, Aachen.

http://dx.doi.org/10.18420/muc2016-ws11-0000

Mobile phone users not only generate data from surfing the web, but also generate location data, app data, and health data. Sensors are ubiquitous and their data needs to be integrated. In face of the Internet of Things, Industrial Internet, or Industry 4.0 and increasingly pervasive digitalization a tsunami of data will revolutionize our everyday lives. Can we use Big Data to tackle global challenges?

Various approaches have been developed to grasp Big Data. *Machine Learning* addresses Big Data by using algorithmic approaches to tackle the sheer size and complexity of data. On the other hand, *Visual Analytics* is a field that tries to combine information visualization – the science of visually displaying quantitative information – with nearby fields, such as knowledge discovery, cognitive and perceptual sciences, statistical analysis. Bringing those two approaches together is the aim of Human-Computer Interaction and Knowledge Discovery in Databases (HCI-KDD, cf. Holzinger, 2013). The overall aim is to support decision-making on the basis of data. Or, how do we get from large amounts of data from the digital world into actionable knowledge in the mental world?

Many of the hard questions have been approached, yet remain partially unanswered. Where is the place of the Human in the Loop (Holzinger 2016)? How do we design interfaces that support the users in making decisions (Sedlmair et al. 2012)? How do we technically create visualizations that represent hard scientific problems (Childs et al. 2013)? How much does a visualization tool need to be tailored the specific problem, how much generalization is possible (Wickham, 2010)? What insights can be drawn from a specific visualization, and by whom (Yi et al. 2008)?

This article reflects on these questions and its implications and elaborates on the meaning of human-factors research for these questions. It furthermore refers to articles with research agendas in their given field that elaborate on these questions.

2 Managing Information Complexity

Managing large amounts of data can be a critical advantage for business in the age of Industrie 4.0 and the Internet of Things. Corporate information management or Enterprise Resource Planning (ERP) systems help decision makers to steer business externally and internally. These systems may encapsulate and aggregate key performance indicators into actionable knowledge or helpful information. Still, these systems rely on machine learning or data analysis techniques not immune to faulty outcomes. When data, algorithm output and personal knowledge are in dissonance with reality bad decisions can be made. How individuals deal with this dissonance – do they favor their knowledge over seemingly perfect algorithms – is highly individual and thus of interest to human factors research. What conditions are necessary for decision makers to adequately judge the information from an ERP System must be understood, before automated decision making or monitoring of cyber-physical systems may become widely accepted. The algorithm-human hybrid decision-making depends on understanding interactions of system, user and context (Brauner et al. 2016).

Understanding these interactions is not only necessary when steering a corporation, but also when trying to automatize analysis of business-critical data from text sources. Text-mining approaches have come a long way; yet, they still rely on human feedback to adequately map topics from documents in new domains. Integrating this feedback into the algorithm can be used to understand research output in a digitalized organization to maintain a competitive advantage in technology forecasting, for instance. (Thiele et al. 2016)

Visualizing the output of such algorithms is mostly done using graph-based approaches, visualizing topic-document relations or co-authorship networks. Various tools exist that help understand the content and social structure encoded in digital libraries. But graph representations, as ingenious they are, come with a sensitive complexity functions. Networks of a given size can exponentially increase in complexity given not only their relations but also their communities. Understanding graph complexity is necessary to help design informative graph-based visualizations that are rich in information, but also understandable (Abels et al. 2016).

3 Medical Applications

The largest field of application of Big Data approaches is medicine. Medicine suffers from quantity and the unstructured nature of its data, often aggregated from different sensors, with different timings and various levels of aggregations. Medical data is highly patient-dependent as no patient is alike. Detecting anomalies from image data or time-series data is hard for humans, thus the large field of biomedical data mining. Still, in the end the doctor has to make a decision based on actionable knowledge. Doctors must understand the capabilities and limitations of algorithms and make judgments accordingly. Integrating the doctor in the algorithm-based decision-making process is crucial for building trust in big-data solutions and to foster understanding of (rare) diseases. One tool are interactive recommender systems (Holzinger et al. 2016) for medical applications. By allowing decision-making with a doctor-in-the-loop they utilize the knowledge and wisdom of doctors in conjunction with the rigor and conscientiousness of machine-learning algorithms. The question is how to design interfaces and how to incorporate domain knowledge as to maximize support for the experts – in this case: doctors.

On the other side of the table we have patients., who want the best-possible treatment and who want to understand their condition. The informed patient wants to know about their health status regularly and keep an eye on their health parameters. One problem is, that the need for information increases with age, because chronic diseases in particular require monitoring of health parameters, which increase with age. But age does not only increase the need for health surveillance, it also increases the diversity in the present population, regarding perceptual and cognitive capabilities. Therefore it is particularly necessary to address patient diversity when visualizing health data to a patient (Theis et al. 2016). Only human factors research and age-dependent health visualization can help leverage the advantages of data-driven health.

4 Uncertainty and Trust

Independent of their usage context, information visualization uses data from sources that are necessarily uncertain. While the data may be discrete and often correct, probabilistic algorithms and the open-world assumption remind us that uncertainty is always present. The more pressing question is how does the visual representation of data incorporate the uncertainty and how can a user communicate their own uncertainty to a computer (Seipp et al. 2016). Scientific publications have an established modus operandi in reporting uncertainty. Disputable as they may be, publications based on quantitative data are seldom found without a p-value or better confidence intervals, communicating the uncertainty in the data. Some of these measures are also found in information visualization or visual analytics, but research is still necessary when trying to visualize where uncertainty comes from (i.e. from the model, data, etc.). Furthermore communicating uncertainty to the computer has hardly been investigated so far. This uncertainty can stem from differences in domain knowledge, experience with visualizations, or expertise in using a system at hand. Human-factors research should address these differences and their implications on human-computer interaction in both directions (Seipp et al. 2016).

Besides uncertainty of data, how visualizations are crafted may impact their utility and their credibility. With rising distrust in elites, authorities and science, it becomes important to investigate the trustworthiness of scientific visualizations. Exploratory research hints to the fact that design-factors of visualizations directly influence credibility (Platte et al. 2016). In particular, the amount of presented visual information may follow a U-shaped curve, causing distrust in visualizations that focus on minimalistic ink usage and information reduction on one end of the spectrum. Here, viewers wonder about what might be (purposefully) hidden from them. On the other end of the spectrum, adding too much ink, e.g. by adding explanatory information, may be perceived as maliciously manipulative, if the visualization hints to a conclusion that is too obvious. Finding the sweet spot of data-to-ink ratio may depend on domain knowledge, personality factors and education and is thus of interest to human factors research.

5 Summary

In order to successfully manage the information age, it is necessary to include human factors research in all fields of data and information applications. In particular, the interfaces between humans and computers, and their consequences must be investigated from a human factors perspective, when moving to the domain of information visualization and decision support systems. Differences in demographics, personality, domain knowledge, experience with visualizations, data and algorithms may all play a role in understanding how to create solutions that leverage data and information for human benefits, be it in medicine, industry or science. Even when all aspects of perceptual and cognitive diversity were addressed, emotional aspects in decision-making would be a core aspect that needed a human factors perspective. Understanding how uncertainty, trust and risk affect decision-making and thus information visualization can address them, is crucial for a holistic and successful approach to Big Data in the future.

Acknowledgements

The authors thank the German Research Council DFG for the friendly support of the research in the excellence cluster „Integrative Production Technology in High Wage Countries“.

Literaturverzeichnis

Childs, H., Geveci, B., Schroeder, W., Meredith, J., Moreland, K., Sewell, C., ... & Bethel, E. W. (2013). Research challenges for visualization software. *Computer*, (5), 34-42.

Holzinger, A. (2013). Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together? In *Availability, Reliability, and security in Information Systems and HCI* (pp. 319-328). Springer Berlin Heidelberg.

Holzinger, A. (2016). *Interactive machine learning for health informatics: when do we need the human-in-the-loop?. Brain Informatics,* 1-13.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). *Big data. The management revolution. Harvard Bus Rev*, 90(10), 61-67.

Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2431-2440.

Wickham, H. (2010). A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, *19*(1), 3-28.

Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A. (2008). 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* (p. 4). ACM.

Abels, S., Greven, C., Calero Valdez, A., Schroeder, U., Zielfe, M. (2016). Graph Complexity in Visual Recommender Systems for Scientific Literature. In Mensch und Computer 2016 – Workshopband

Brauner, P., Calero Valdez, A., Philipsen, R., Ziefle, M. (2016). On Studying Human Factors in Complex Cyber-Physical Systems. In Mensch und Computer 2016 – Workshopband

Holzinger, A., Calero Valdez, A., Ziefle, M. (2016). Towards Interactive Recommender Systems with the Doctor-in-the-Loop. In Mensch und Computer 2016 – Workshopband

Platte, L. Freitag, S., Schubert, L., Molis, J., Calero Valdez, A., Ziefle, M. (2016). Trustworthiness of visualizations of mobility-induced CO2 emissions. In Mensch und Computer 2016 – Workshopband

Seipp, K., Ochoa, X., Gutiérrez, F., Verbert, K. (2016). A Research Agenda for Managing Uncertainty in Visual Analytics. In Mensch und Computer 2016 - Workshopband

Theis, S., Bröhl, C., Rasche, P., Wille, M., Schlick, C.M., Mertens, A. (2016). Age-dependent health data visualizations: a research agenda. In Mensch und Computer 2016 – Workshopband

Thiele, T., Sommer, T., Schröder, S., Richert, A., Jeschke, S. (2016). Human-in-the-loop Processes as Enabler for Data Analytics in Digitalized Organizations. In Mensch und Computer 2016 - Workshopband

Autoren

Calero Valdez, André

André Calero Valdez has studied computer science at the RWTH Aachen University and holds a PhD in Psychology also from RWTH Aachen University. He a senior researcher at the Human-Computer Interaction Center of the RWTH Aachen University and visiting professor with the HCI-KDD group in Graz, Austria. His thesis dealt with the topic of user-centered design of small screen devices for diabetes patients. He currently conducts research in the topics of knowledge management, social media, and decision support by visualizations. The aim is to manage complexity of information by applying human-computer interaction principles.

Brauner, Philipp

Philipp is computer scientist, psychologist, and communication scientist and researches human performance in cyber-physical systems in various domains. He is an expert in using serious games, game-based learnings, and gamification to increase performance, motivation, and engagement in interacting with CPS.

Ziefle, Martina

Martina Ziefle holds the chair of communication science and is founding member of the Human-Computer Interaction Center of the RWTH Aachen University. Her research addresses the communication between human-human and human-machine with the research focus on technology acceptance for various technologies with respect to user diversity.

Kuhlen, Torsten Wolfgang

Torsten Kuhlen has been the head oft he Virtual Reality Group at the RWTH Aachen University since 1998 and holds a profosserhip in computer science since 2008. His research foci encompass the analysis of complex numerical simulations in virtual environments, the usage of VR technology in medical simulations and neuropsychology, as well as 3D acoustics in virtual environments.

Sedlmair, Michael

Michael Sedlmair has studied media informatics at the LMU Munich and attained a PhD in collaboration with BMW Group Research and Technology and the LMU Munich in Computer Science. The topic was Visualization of bus communication in vehicles. Currently his research interests are visualization, human-computer interaction, user-centered design and high-dimensional data.