Challenges of Time-oriented Data in Visual Analytics for Healthcare

Wolfgang Aigner* Paolo Federico† Theresia Gschwandtner‡ Silvia Miksch§ Alexander Rind¶

Institute of Software Technology and Interactive Systems
Vienna University of Technology

Abstract

The visual exploration and analysis of time-oriented data in healthcare are important yet challenging tasks. This position paper presents six challenges for Visual Analytics in healthcare: (1) scale and complexity of time-oriented data, (2) intertwining patient condition with treatment processes, (3) scalable analysis from single patients to cohorts, (4) data quality and uncertainty, (5) interaction, user interfaces, and the role of users, and (6) evaluation. Furthermore, it portrays existing and future work by the authors tackling these challenges.

Index Terms: H.5.m [Information Systems]: Information Interfaces And Presentation (e.g., HCI)—Miscellaneous I.3.6 [Computing Methodologies]: Computer Graphics—Methodology and Techniques; J.3 [Computer Applications]: Life and Medical Sciences—Medical information systems

1 Introduction

Utilizing the huge volumes of heterogeneous data resources and collections is one of the greatest challenges of our computerized society. This holds in particular for healthcare, where different user groups are collecting, assessing, exploring, and analyzing such kinds of data and information. Visual Analytics denotes "the science of analytical reasoning facilitated by visual interactive interfaces" [22] and aims to make complex information structures more comprehensible, facilitate new insights, and enable knowledge discovery. Visual Analytics methods focus on the information discovery process exploiting both the computational power of computers and the human's visual information processing capabilities. Therefore, it aims to enable the exploration and the understanding of large and complex data sets intertwining interactive visualization, data analysis, and human-computer interaction.

2 Challenges

In the last years, different articles summarized open problems and main challenges for Visual Analytics (cp. [12, 13, 21, 22, 23]). We surveyed the state-of-the-art of information visualization approaches for exploring and querying Electronic Health Record systems (EHRs) in a recent article [20] and collected visualization methods of time-oriented data and information [2]. According to these references, we illustrate the most important open problems and challenges for Visual Analytics in Healthcare and present possible solutions to some issues.

Scale and Complexity of Time-oriented Data. Usually, the heterogeneous data resources and collections in healthcare are not only large and complex, but also time-oriented. In contrast to other quantitative data dimensions that are usually "flat", time has inherent semantic structures, contains natural cycles and re-occurrences (as for example seasons), but also social (often irregular) cycles, like holidays or school breaks. For example, the time span between check-up examinations of a chronic patient may vary between weeks and years. Therefore, time-oriented data need to be treated differently from other kinds of data and demand appropriate interaction, visual and analytical methods to analyze them.

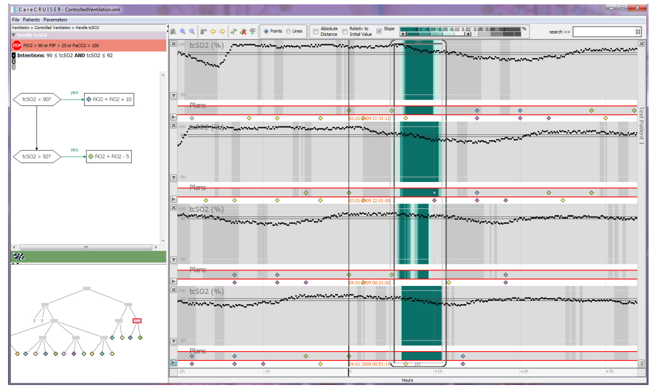

Figure 1: CareCruiser [9]: The temporal view (on the right) arranges patient parameters together with applied clinical actions along a horizontal time axis. In this screenshot the turquoise color marks the falling of the $tcSO_2$ values. Multiple instances of applying the same clinical action to one patient are aligned on a vertical axis.

Intertwining Patient Condition with Treatment Processes. Healthcare data, such as in EHRs, cover not only observations about the patients' condition, but also information about the various treatment actions over time. All these data and information need to be analyzed intertwinedly. The medical staff usually does not examine a single patient parameter, but observes the correlations of multiple parameters to assess the patient's health condition. Moreover, the parameter value at a single point in time is less meaningful than its evolution over time. In particular, the identification of changes in a patient's condition in reaction to applied treatments, demands for an intertwined view. The following tasks are high-level tasks in medical care and require a representation of the patient's parameters (i.e., the health condition of the patient) in tight combination with the applied treatment actions:

1. Monitoring the treatment progress (i.e., which treatment action is being applied at the moment, which treatments have been applied so far, and which actions may be applied in the near future),
2. Monitoring the overall success or failure of applied treatments,
3. Seeing the effects of different treatment actions on the individual patient's condition,
4. Getting a comprehensive picture about the possible reasons for changes of the patient's condition (i.e., the bettering or worsening of single patient parameters),

*e-mail: aigner@ifs.tuwien.ac.at
†e-mail: federico@ifs.tuwien.ac.at
‡e-mail: gschwandtner@ifs.tuwien.ac.at
§e-mail: miksch@ifs.tuwien.ac.at
¶e-mail: rind@ifs.tuwien.ac.at

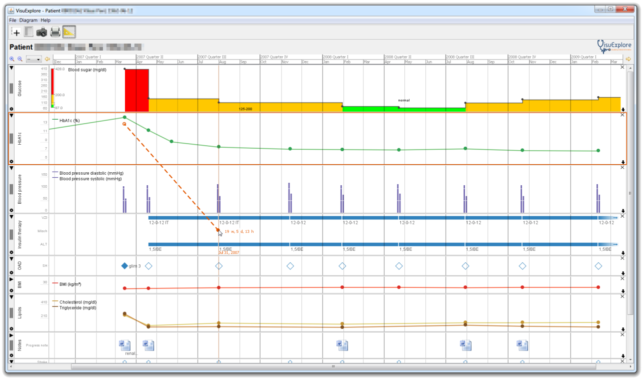

Figure 2: VisuExplore [19]: Overview visualization of a patient's medical history predominantly using well-know and easy to read visual representations (e.g., line plots).

5. Identifying sub-optimal treatment choices, and thus,
6. Optimizing treatment for individual patients.

Considering the information needs of different user groups, the following visualization prototypes all tackle specific aspects of these tasks.

CareVis [1] is an interactive visualization that displays patient data in combination with computer-interpretable medical guidelines and protocols (CGP), which capture the complex structural and temporal constraints of applied and future treatment. CareCruiser [9] is a conceptual extension of CareVis with a special focus on the exploration of the effects of clinical actions on a patient's condition. It provides several features to support a step-wise interactive exploration: (1) aligning clinical actions, (2) color-coding curve events, (3) filtering color-coded information, and (4) a focus & context window for the detection of patterns of effects (Figure 1). VisuExplore [19] is more powerful regarding patient data but less regarding the structural and temporal constraints of treatment. It can display various aspects of an EHR by supporting different visualization methods in parallel panels along a common time axis. For example, in Figure 2, medical test values are represented by line plots, bar charts, and a step chart with color-coded qualitative abstractions. Treatment performed over a period of time is shown in a timeline chart through horizontal bars.

Scalable Analysis from Single Patients to Cohorts. Healthcare requires scalable visualization and analysis methods. Besides the complexity and the scale of the time dimension and the multivariate nature of healthcare data, an additional dimension to consider is the number of patients to be analyzed simultaneously. Indeed, while a system focusing on the analysis of a single patient might be sufficient to provide appropriate care tailored to the needs of that specific patient, multiple-patients systems can be useful to compare the response of diverse patients, to follow the development of an entire cohort, or to assess the effectiveness of a therapy on a larger scale.

We have proposed different solutions for the visual analysis of multiple patients. CareCruiser [9] (Figure 1) enables the exploration of two or more patients, providing collapsible facets each showing the evolution of one patient along the time axis; to support a better comparison, the data can be interactively aligned using a relative time (e.g., calendar date, time since start of therapy, or time since any other event). TimeRider [18] and Gravi++ [11] exploit animation and traces to show the evolution of multiple patients. TimeRider (Figure 3) enables bi-variate analysis of cohort trends by the means of animated scatter plots: marks representing patients are laid out according to two categorical or numerical axes and animated to show their temporal evolution; data wear is encoded to transparency, to take into account different sampling rates.

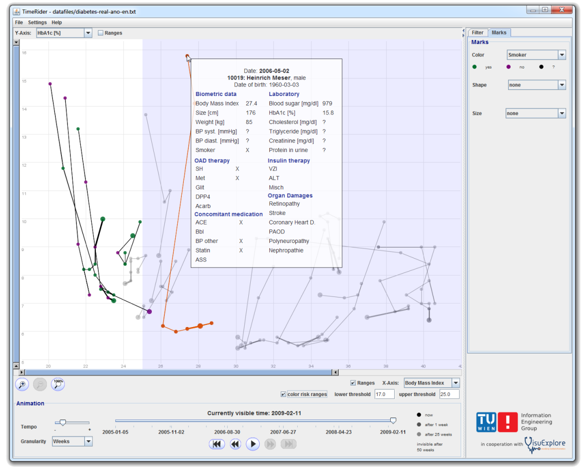

Figure 3: TimeRider [18]: Animated scatter plot for bi-variate analysis of patient cohort trends. Optional traces show the complete trajectory of the patients. The blue background denotes elevated parameter values ($BMI > 25$).

Gravi++ (Figure 4) enables multi-variate analysis: different patients are spatially clustered by a dynamic spring-based layout taking into account several variables. An analogous incidence-model could be used to visualize patients' cohorts as dynamic networks. In ViENA [7] we have integrated different static visualizations for dynamic networks, namely juxtaposition (small multiples), superimposition, and 2.5D views (Figure 5). The benefits and limitations of animation and static visualizations of patients' parameters need further investigations also concerning the different medical users.

Knowledge-based temporal abstractions, besides supporting specific user needs in the medical domain by combining quantitative and qualitative aspects, also enable more compact visualizations. With Midgaard [6], a visualization that combines raw data and abstractions as well as a semantic zoom changing the level of abstraction has been introduced (Figure 6). Moreover, we have evaluated its effectiveness in supporting tasks involving parameters of single patients [4]. Such compact visualizations based on temporal abstractions can be useful when dealing with multiple patients as a mean to optimize the display space occupancy, but their application in the case of multiple variables and multiple patients should be researched further. Furthermore, a promising research topic is the development of context-based temporal abstractions that pursue a closer interaction with knowledge, adapting to the patient's context and reacting dynamically to its modifications.

Data Quality and Uncertainty. A central issue in Visual Analytics is to avoid misinterpretation by the analysts. However, in real-life data there are several issues that may lead to misinterpretation or wrong results, such as missing data, uncertain data, ambiguous data, or simply wrong data. Especially in the discipline of healthcare, data sets may contain an unavoidable amount of uncertainty, errors, and ambiguity. To assure the reliability of any data analysis step, quality problems within the data set have to be detected and – if possible – resolved first. Several taxonomies of general data quality problems exist, but they do not consider the very special characteristics of time (in healthcare, data sets are highly time-oriented). To this end, we have provided a taxonomy of time-oriented data quality problems [10]. On the one hand, it gives a unified view on the various existing taxonomies of general data quality problems. On the other hand, it provides an important reference when formulating quality checks of time-oriented data.

However, there may be data issues that cannot be corrected, such

Figure 4: Gravi++ [11]: A spring-based layout is used for spatial clustering by multiple parameters, which are represented by six squares in the screenshot. Animation and traces show evolution over time.

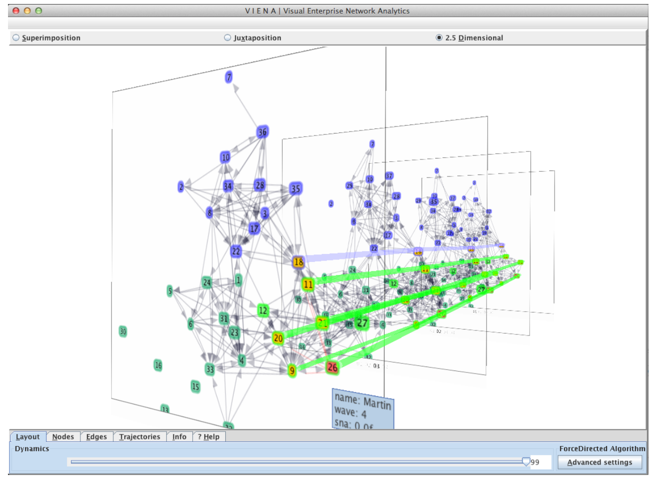

Figure 5: ViENA [7]: 2.5D visualization of dynamic networks. Traces represent change of network metric for person nodes over time.

as an uncertain starting time of a future event. These issues have to be communicated appropriately to the user in order to ensure an informed interpretation of the data at hand. PlanningLines [3], for instance, use novel glyphs to visualize temporal uncertainties. The glyph visually communicates the earliest and latest possible starting time of a task, the earliest and latest possible ending time, as well as the minimum and the maximum duration of the task. It was designed to represent complex time annotations for CGPs.

Interaction, User Interfaces, and the Role of Users. Large and complex data sets cannot be visualized and analyzed as a whole at once. Exploration of these data sets is an interactive, multi-step process that involves trial-and-error, human judgment, and exchange with colleagues. Therefore, task-specific interaction methods and user interfaces are required. Furthermore, Visual Analytics methods need to account for the different backgrounds and usage contexts of the user groups involved in healthcare. While physicians or nurses are driven by tight schedules and frequent interruptions towards simple interfaces that deliver overview at a glance, clinical researchers and quality analysts need flexibility and support for their reasoning process. Non-professionals such as patients, family member carers, or other intermittent users play an important role in healthcare, but need to be addressed more specifically by Visual Analytics methods.

We follow a user/data/task-centered design approach in all our application-oriented projects and we develop task-specific and user-specific interaction methods and user interfaces. For example, CareCruiser [9] is tailored for CGP-based care as it allows for an active investigation of the development of the patient's condition and the detection of effects of applied CGPs. VisuExplore [19] relies on well-know and easy to read visual representation techniques such as line plots and timeline charts to provide a clear and unambiguous overview of a patient's medical history. Furthermore, it allows personalization of the user interface either by interaction or through a configuration file.

There is a need for well-defined process models in Visual Analytics, in order to better understand the analytical reasoning process and develop more suitable Visual Analytics methods. We empirically analyzed interaction logs collected from user studies of Gravi++ and VisuExplore and identified common interaction patterns and transition probabilities [17]. In related work, we describe a Visual Analytics process that uses the structure of time to build hypotheses and statistical models on time-series data [14].

Evaluation. The Visual Analytics process is complex, comprised of multiple steps of computation and human reasoning, and produces outcomes that are hard to measure. Thus, it is very difficult to assess the quality and effectiveness of Visual Analytics methods, in particular in an interdisciplinary domain like healthcare. Nevertheless, evaluation is essential both for adoption in clinical practice and advancing Visual Analytics for Healthcare as a scientific community.

Evaluation methods can be categorized by the threats they address in the design process [16]. Evaluation against certain threats requires the involvement of domain experts (e.g., physicians), but their tight schedules make it hard to recruit more than a few subjects. The combination of different methods can alleviate these problems and strengthen the evidence on Visual Analytics methods (e.g., Gravi++ [11]).

To compare and assess various Visual Analytics solutions, large benchmark data sets of de-identified patient records, relevant tasks, and gold standard solutions would be necessary. On the other hand, if a user-centered design process is followed and concrete tasks and data of the involved users are addressed, established categorizations can be used to make the results better comparable. For that purpose, we regularly apply the task framework by Andrienko and Andrienko [5], the user intents by Yi et al. [24], and the heuristics by Forsell and Johansson [8]. Furthermore, we have proposed a categorization for time-oriented data [2] and a task framework that is extended along the structure of time [15].

Finally, many steps of an evaluation study such as task display, time keeping, and data collection can be automated. We are working on a general evaluation framework that can be plugged into Visual Analytics prototypes. It has been tested successfully in several user studies.

Other Open Problems and Challenges. We are aware that the above list does not cover all issues, for example, we did not elaborate about infrastructures, hardware, display and interaction devices, data streams, patient safety, data security, personalization, or privacy. However, we aim to address the most important issues specific to healthcare, first.

3 Outlook

Our research group will tackle these challenges in two current research projects and future work. In the course of the Laura Bassi Centre of Expertise *CVAST*,[1] we aim to develop novel, user-

[1] `http://www.cvast.tuwien.ac.at/`, cited Aug 31, 2012.

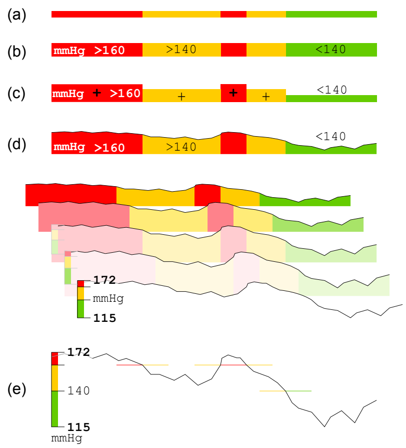

Figure 6: Midgaard [6]: Visualization technique for numerical variables with semantic zoom to one of five levels of detail: (a) colored background, (b) colored background with labels, (c) colored bars, (d) colored area charts, and (e) augmented line charts.

oriented, and task-specific Visual Analytics methods that foster new insights and enable knowledge discovery. Through our participation in the *MobiGuide* project,[2] we aim to design and develop Visual Analytics methods for the patients' data and the guideline processes, focusing on and their compliance and modifications over time, also addressing uncertainty and incompleteness.

Acknowledgements

The Laura Bassi Centre of Expertise CVAST is funded by the Austrian Federal Ministry of Economy, Family and Youth (project number: 822746). MobiGuide is funded by the European Commission under the 7th Framework Program (grant number: 287811).

References

[1] W. Aigner and S. Miksch. CareVis: Integrated visualization of computerized protocols and temporal patient data. *Artificial Intelligence in Medicine*, 37(3):203–218, 2006.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, London, 2011.

[3] W. Aigner, S. Miksch, B. Thurnher, and S. Biffl. PlanningLines: Novel glyphs for representing temporal uncertainties and their evaluation. In *Proc. 9th Int. Conf. Information Visualisation (IV 2005)*, pages 457–463. IEEE, 2005.

[4] W. Aigner, A. Rind, and S. Hoffmann. Comparative evaluation of an interactive time-series visualization that combines quantitative data with qualitative abstractions. *Computer Graphics Forum*, 31(3):995–1004, 2012.

[5] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, Berlin, 2006.

[6] R. Bade, S. Schlechtweg, and S. Miksch. Connecting time-oriented data and information to a coherent interactive visualization. In *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI)*, pages 105–112, 2004.

[7] P. Federico, W. Aigner, S. Miksch, F. Windhager, and L. Zenk. A visual analytics approach to dynamic social networks. In *Proc. 11th Int. Conf. Knowledge Management and Knowledge Technologies (i-KNOW '11)*, pages 47:1–47:8. ACM, 2011.

[8] C. Forsell and J. Johansson. An heuristic set for evaluation in information visualization. In G. Santucci, editor, *Proc. Int. Conf. Advanced Visual Interfaces (AVI 2010)*, pages 199–206. ACM, 2010.

[9] T. Gschwandtner, W. Aigner, K. Kaiser, S. Miksch, and A. Seyfang. CareCruiser: exploring and visualizing plans, events, and effects interactively. In *Proc. IEEE Pacific Visualization Symp. (PacificVis 2011)*, pages 43–50, 2011.

[10] T. Gschwandtner, J. Gärtner, W. Aigner, and S. Miksch. A taxonomy of dirty time-oriented data. In G. Quirchmayr, J. Basl, I. You, L. Xu, and E. Weippl, editors, *Multidisciplinary Research and Practice for Information Systems, Proc. CD-ARES 2012*, LNCS 7465, pages 58–72. Springer, 2012.

[11] K. Hinum, S. Miksch, W. Aigner, S. Ohmann, C. Popow, M. Pohl, and M. Rester. Gravi++: Interactive information visualization to explore highly structured temporal data. *Journal of Universal Computer Science*, 11(11):1792–1805, 2005.

[12] D. Keim, G. Andrienko, J. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. T. Stasko, J. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, LNCS 4950, pages 154–175. Springer, Berlin, 2008.

[13] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. *Mastering The Information Age – Solving Problems with Visual Analytics*. Eurographics, Goslar, Germany, 2010.

[14] T. Lammarsch, W. Aigner, A. Bertone, S. Miksch, and A. Rind. Towards a concept how the structure of time can support the visual analytics process. In S. Miksch and G. Santucci, editors, *Proc. Int. Workshop on Visual Analytics (EuroVA 2011) in conjunction with EuroVis 2011*, pages 9–12. Eurographics, 2011.

[15] T. Lammarsch, A. Rind, W. Aigner, and S. Miksch. Developing an extended task framework for exploratory data analysis along the structure of time. In K. Matkovic and G. Santucci, editors, *Proc. Int. EuroVis Workshop on Visual Analytics (EuroVA 2012)*, pages 31–35. Eurographics, 2012.

[16] T. Munzner. A nested process model for visualization design and validation. *IEEE Trans. Visualization and Computer Graphics*, 15(6):921–928, 2009.

[17] M. Pohl, S. Wiltner, S. Miksch, W. Aigner, and A. Rind. Analysing interactivity in information visualisation. *KI – Künstliche Intelligenz*, 26:151–159, May 2012.

[18] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, F. Drexler, B. Neubauer, and N. Suchy. Visually exploring multivariate trends in patient cohorts using animated scatter plots. In M. M. Robertson, editor, *Ergonomics and Health Aspects of Work with Computers, Proc. Int. Conf. held as part of HCI International 2011*, LNCS 6779, pages 139–148. Springer, 2011.

[19] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, T. Turic, and F. Drexler. Visual exploration of time-oriented patient data for chronic diseases: Design study and evaluation. In A. Holzinger and K. Simonic, editors, *Information Quality in e-Health, Proc. USAB 2011*, LNCS 7058, pages 301–320. Springer, 2011.

[20] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman. Interactive information visualization to explore and query electronic health records. *Foundations and Trends in Human-Computer Interaction*, 2012. In review.

[21] J. Thomas and J. Kielman. Challenges for visual analytics. *Information Visualization*, 8(4):309–314, 2009.

[22] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE, 2005.

[23] P. C. Wong, H.-W. Shen, C. R. Johnson, C. Chen, and R. B. Ross. The top 10 challenges in extreme-scale visual analytics. *IEEE Computer Graphics and Applications*, 32(4):63–67, Aug. 2012.

[24] J. S. Yi, Y. A. Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Visualization and Computer Graphics*, 13(6):1224–1231, 2007.

[2] `http://www.mobiguide-project.eu/`, cited Aug 31, 2012.