International Workshop on Visual Analytics (2011)
S. Miksch and G. Santucci (Editors)

Towards a Concept how the Structure of Time can Support the Visual Analytics Process

T. Lammarsch[1], W. Aigner[1,2], A. Bertone[2], S. Miksch[1], and A. Rind[1]

[1]Institute of Software Technology and Interactive Systems (ISIS), Vienna University of Technology, Austria
[2]Department of Information and Knowledge Engineering (ike), Danube University Krems, Austria

Abstract

The primary goal of Visual Analytics (VA) is the close intertwinedness of human reasoning and automated methods. An important task for this goal is formulating a description for such a VA process. We propose the design of a VA process description that uses the inherent structure contained in time-oriented data as a way to improve the integration of human reasoning. This structure can, for example, be seen in the calendar aspect of time being composed of smaller granularities, like years and seasons. Domain experts strongly consider this structure in their reasoning, so VA needs to consider it, too.

Categories and Subject Descriptors (according to ACM CCS): Information Systems [H.1.1]: Models and Principles—Systems and Information Theory; Computing Methodologies [I.m]: Miscellaneous—

1. Introduction

Visual Analytics (VA) is defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [TC05]. The combination of automated reasoning methods and visual interfaces is a necessary step, but only an intermediate one to the combination of automated and human analytical reasoning. Current descriptions of the VA process are either human-focused and difficult to apply for developing VA systems [TC05, PC05] or process-oriented frameworks that do not place much consideration into human reasoning [KMS*08, BL09]. To bridge the gap between human reasoning and automated methods, we propose to include specific characteristics of certain kinds of data. For many applications of time-oriented data, the structure of time (see Section 2 for a discussion of this term in state-of-the-art research) is of key importance in the reasoning of human users [SML*08, SML*09]. Our goal is to make the structure of time an integrated part of the VA process. To pioneer that development in this paper, we (1) propose a VA process in a way suitable for our goal and (2) present a concept for organizing data according to the structure of time.

2. Related Work

We present existing descriptions of the VA process and important research regarding data concepts for the structure of time. Most of today's VA research is based on the work of Thomas and Cook [TC05]. They describe the analytical reasoning process based on a sense-making loop. The work is overall an advancement of a description by Pirolli and Card [PC05] based on human actions in intelligence analysis. This description has a focus on the early steps of the process, finding hypotheses in a vast amount of seemingly unrelated data, but less detail on the later steps, the validation of hypotheses, building of models, and forecasting of data. A description which contains several sub-processes is presented by Green et al. [GRF09]. Most sub-processes deal with the same tasks as the description by Pirolli and Card, one of them handles the generation and analysis of hypotheses. Keim et al. [KMS*08] provide a process description as an integrated view on visualizations and automated analysis. The heart of the description is a diagram showing the VA process with four states: datasets, visualizations, hypotheses, and insights. The idea of the description is to iterate over the states while increasing the amount of existing hypotheses and visualizations, generating an increasing number of insights. Bertini and Lalanne [BL09] focus on the analysis of methods for the integration of a human-centered approach to knowledge discovery and machine-driven Data Mining. They do consider Data Mining as a transition while they consider visualizations as a state—other approaches usually consider them as equivalents. Furthermore, Bertini and Lalanne state the important fact that while it is possible for humans to form what they call a mental schema based on visualizations with sensible defaults, it is essential for Data Mining to have this mental schema before performing the mining algorithms because proper parametrization is much more important. The process descriptions are kept independent of the types of data. To develop possibilities focused on the structure of time, we also have to take a look at concepts for structurizing time-oriented data (according to aspects which have been, for example, described by Aigner et al. [AMST11]). Jensen and Snodgrass introduced the BCDM [JS96] and applied the important concepts of chronons and intervals, but focused on database transactions which are not of central importance in VA systems. The HMAP by Combi and Pozzi [CP01] complements the interval primitive with instants and durations. It also applies the calendar model of granularities, which has been described by Bettini et al. [BJW00]. Those aspects are implemented for ontologies by Hobbs and Pan [HP04], as well as a definition of temporal before/after relations.

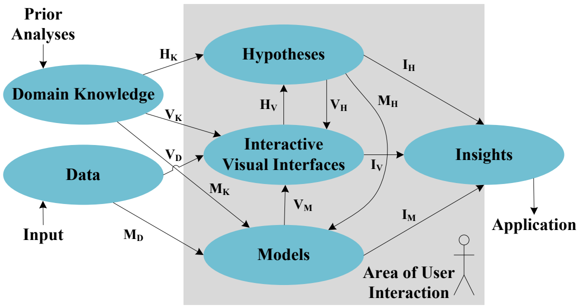

Figure 1: *Our VA process, based on [KMS*08] and [BL09].* H_K*: Take hypotheses from domain knowledge.* V_K*: Visualize domain knowledge directly.* M_K*: Take models from domain knowledge.* V_D*: Visualize data.* M_D*: Generate models from data.* H_V*: Build hypotheses based on visualizations.* V_M*: Visualize models.* V_H*: Visualize hypotheses.* M_H*: Validate hypotheses to form models.* I_H*: Gain insights from hypotheses.* I_M*: Gain insights from models.* I_V*: Gain insights from visualizations.*

3. Our Process Description

For our process description, we take several aspects from the ones by Keim et al. [KMS*08] as well as Bertini and Lalanne [BL09] (see Figure 1). Our description is not a state machine. Rather, process elements are generated and accumulated over the course of the process. Users can interact with anything inside the grey area (including the arrows that lead inside and outside), mainly by using the interactive visual interfaces that double as process element as well as interface. **Data** are real-world values collected by prior processes. **Domain Knowledge** consists of hypotheses and models from prior processes that are considered. **Interactive Visual Interfaces** are any representation of data that are intended for transferring them from automated systems to human users through the sense of sight. Visual interfaces can also have the ability to transfer data from users to automated systems through user interaction. **Models** are representations of a system of entities, phenomena, or processes. In many present process descriptions and essays about the scientific method, hypotheses are a subclass of a model. We narrow the definition and consider only those results a model that are validated by comparison to existing data. Therefore, results not validated are called a hypotheses. Still, models cannot be considered “true” for sure, because they are only correct for the current state of data collection. In automated systems they often stem from Data Mining methods, but in VA, an important source is the validation of hypotheses. **Hypotheses** consist either of a suggested explanation for an observable problem or of a reasoned proposal predicting a possible causal correlation among multiple phenomena. We distinguish between (1) hypotheses formulated in common language or thoughts of users and (2) externalized hypotheses formulated in a way that can be used by automated systems (in the following, we will consider externalization always in that context). To keep the process description manageable, we do not introduce two process elements, as both variants have the same content. Externalizing hypotheses is of grave importance for the VA process because it is the key to their validation. Validated hyotheses become models. In theory, methods like fuzzy logic can also generate hypotheses without human users, but as their results are usually validated automatically, we have omitted additional transitions for the sake of simplicity. Hypotheses can also be visualized, often in form of annotations. It is possible for VA systems to use interactive visual methods to externalize hypotheseses, like Fuchs et al. [FWG09] do with linking and brushing. **Insights** are understandings of coherence gained by users that were not clear before the process. Insights can cause users to consider certain data, hypotheses or models to guide their further actions or analyses.

4. A Data Concept Using the Structure of Time

We need to handle time-oriented data according to the structure of time as well as according to the VA process model. Hypotheses and models are handled similar to data. We expect methods from VA to be easier applicable and modularly replaceable if the concepts for data, hypotheses, and models are closely related, yet clearly distinguished. We adapt the concept of temporal objects from Aigner [Aig06]. A temporal object is a pair <Object, Temporal Element>, where, according to Aigner, the object can be anything, including another temporal object. We use specific kinds of VA process temporal objects for data, a hypothesis, or a model. VA Process Temporal Object := <VA Process Object, Temporal Element> with VA Process Object := <Type, Further Information, Object>. Type can be data, hypothesis, and model. Further information for data is the source, for hypotheses it is the user who made it, for models it is the applicability based on a certain kind of data for which it has been validated. The structure of time has been defined in various ways, most of which acknowledge several aspects. Therefore, the temporal element of a temporal object needs a structure that can represent those. We rely on the aspects listed by Aigner et al. [AMST11]: **Scope**: The temporal element of a temporal object is formulated as instant, span, or interval. Intervals can be defined based on two instants or on one instant and one span (see Figure 2, without detailing the VA process additions in the objects). The definition of scope becomes more complex in the context of the granularities (see below). **Viewpoints:** We explicitely allow the combination of consecutive data (see Figure 3). A pattern based on several events found by Data Mining methods, like the MuTinY approach by Bertone et al. [BLT*10], can be handled by our concept as a combined temporal object from several smaller temporal objects (the events). This is usually an ordered relation. We expand this relation with optional additional data to objects that are combined in that way as temporal objects. These data describe the relation, e.g. “cause” and “effect’. By doing this, branching time can be handled. By labeling “parallel”, multiple viewpoints can be handled. In most real situations, these additional data have to be provided by users, therefore the resulting temporal objects are hypotheses. **Granularities/Structure:** Granularities are mappings of the units in the discrete time domain (called chronons) to larger units. In the temporal objects concept without granularities, instants are given as absolute chronons and spans are given as an amount of chronons. Using granularities instead has various advantages. Users often think in granularities [SML*08,SML*09]. Therefore, it is easier for them to externalize hypotheses when using them. Generating visualizations using granularities is easier and more effective when a granularitiy-based datal model lies beneath [Lam10], the resulting visualizations are more powerful. We apply granularities by allowing an instant to be a chronon or a granule. A span can be a number of chronons or a number of granules. When somewhere in the VA process a temporal element is asked for, it is possible to pose this question on any granularity level and automatical adaption is performed. A granule is an instant in its own granularity. In granularities below or on the discrete time domain, it is an interval. Therefore, when asking for a granule combined with a finer granularity, the result will be an interval, not an instant. Humans apply granules also as spans. Depending on linguistic context, it is possible that an instant shifted by a span becomes an interval. For example, in strict logics, “in three days” would mean the exact same instant three days later, but it is often used as the whole interval of the day (from midnight to midnight) that contains the shifted instant. Granule labels are not necessarily explicit. For example, “Sunday” recurs every seven days. Therefore, “day of week” is called a “cyclic granularity”. By combining non-temporal elements with “Sunday” as temporal element, it is possible to declare that these aspects occur every Sunday. It is more common to use this possibility for hypotheses and models. By using cyclic granularities, the structure aspect can be handled in our concept. **Scale:** We base our time concept on the assumption that chronons are always discrete. The non-temporal elements also need consideration. For data, it is sufficient to have absolute values of some kind here that are not related to time by itself. In order to deal with models and hypotheses, we have to deal with value ranges and other more complicated data types as non-temporal elements of the objects. Furthermore, the relations between temporal objects need more complex labelling possibilities. For example, the relation “greater than” allows for modelling that the value of one temporal object is greater than the value of another temporal object.

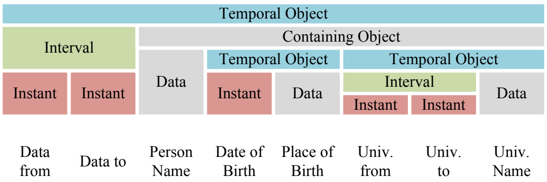

Figure 2: *Temporal objects can be formed according to rows of a data table.*

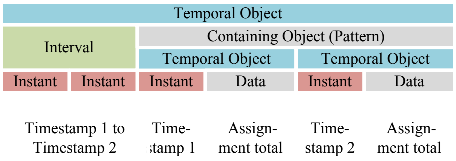

Figure 3: *Temporal objects can also be formed according to columns of a data table.*

5. Application on a Real World Example

We provide a theoretical application of our VA process to a real-world example from prior work (see Figure 4). By looking at a GROOVE visualization based on a dataset of police assignments [LAB*09, p. 7], and based on domain knowledge, a user forms the hypothesis that the 6th and 7th hours of each day have a higher number of assignments than the other hours. This hypothesis can be externalized by forming a granularity called “rush hour” that consists of those two hours and building two temporal objects (“rush hour” and “non rush hour” as labels) for each day. Based on this externalized hypothesis, an automated method can aggregate the chronons from the dataset and count the fraction of the days for which it is true, with the result being a model. The user can gain insights from this model. Using the same dataset, a MuTIny [BLT*10] implementation might produce the result that 75% of the first five minutes of an hour with above average assignments are followed by 55 minutes of below average assignments. This statement, a model, is difficult to comprehend, so the user can visualize it to grasp the meaning. After that, the user might form the hypotheses that this pure consecutive temporal object has a causal connection, labeling the relation between the temporal object. By integrating the structure of time in the concept, users can externalize their hypotheses much more easily and have a better insight in the working of the system.

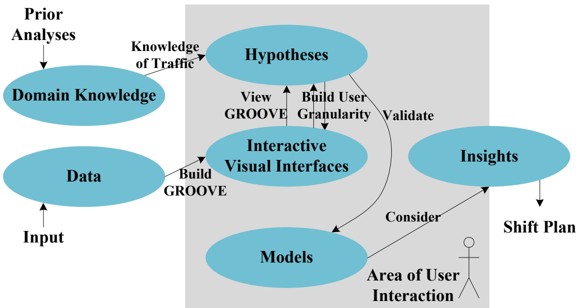

Figure 4: *Our VA process applied to a real-world example. Only the transitions existing in the first example from Section 5 are shown.*

6. Conclusion and Future Work

Based on an overview of the state of the art in VA process descriptions and the state of the art in concepts for time-oriented data, we have presented our unified and expanded process model. It makes a clear distinction between hypotheses and models and therefore the important step of hypotheses validation. Generating models not only through automated methods but also by human reasoning is an important aspect of VA. Our process model has been developed to support the integration of the structure of time, but the state presented in this paper is general enough to be applied for other use cases, too. Furthermore, we have presented a concept for time-oriented data that can handle the several aspects of the structure of time in regards to data, hypotheses, and models. We have also shown how the time concept can be applied to traverse the process elements. Our next steps will be (1) further integration of the process description and the data concept into a reference for the VA process of time-oriented data that is fully focused on the structure of time, (2) defining a scenario of use cases related to the structure of time and enlist users with real-world problems (3) finding suitable methods for performing the various steps of the VA process for our use cases (4) building an actual data structure to hold our concept (5) provide a prototypical implementation (6) as a validation, test the application of our method to the use cases by the users. We expect our approach to provide valuable improvements of the application of VA to time-oriented data according to the structure of time.

Acknowledgments This work was supported by the Austrian Science Fund (FWF). Project number: P22883.

References

[Aig06] AIGNER W.: *Visualization of Time and Time-Oriented Information: Challenges and Conceptual Design*. Ph.D. thesis, Vienna University of Technology, Feb. 2006. Supervisors: Silvia Miksch (Vienna University of Technology), Heidrun Schumann (University of Rostock). 2

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*. Springer, 2011. forthcoming. 2, 3

[BJW00] BETTINI C., JAJODIA S., WANG S.: *Time Granularities in Databases, Data Mining and Temporal Reasoning*. Springer-Verlag New York, Secaucus, NJ, USA, 2000. 2

[BL09] BERTINI E., LALANNE D.: Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. In *Proc. of VAKD09* (2009), ACM, pp. 12–20. 1, 2

[BLT*10] BERTONE A., LAMMARSCH T., TURIC T., AIGNER W., MIKSCH S., GAERTNER J.: MuTIny: A Multi-Time Interval Pattern Discovery Approach To Preserve The Temporal Information In Between. In *Proc. of ECDM10* (2010). 3

[CP01] COMBI C., POZZI G.: HMAP–A temporal data model managing intervals with different granularities and indeterminacy from natural language sentences. *The VLDB Journal 9*, 4 (2001), 294–311. 2

[FWG09] FUCHS R., WASER J., GRÖLLER E.: Visual human+machine learning. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1327–1334. 2

[GRF09] GREEN T., RIBARSKY W., FISHER B.: Building and Applying a Human Cognition Model for Visual Analytics. *Information Visualization 8*, 1 (2009), 1–13. 1

[HP04] HOBBS J., PAN F.: An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing (TALIP) 3*, 1 (2004), 66–85. 2

[JS96] JENSEN C., SNODGRASS R.: Semantics of time-varying information. *Information Systems 21*, 4 (1996), 311–352. 2

[KMS*08] KEIM D., MANSMANN F., SCHNEIDEWIND J., THOMAS J., ZIEGLER H.: Visual Analytics: Scope and Challenges. *LNCS 4404* (2008), 76–90. 1, 2

[LAB*09] LAMMARSCH T., AIGNER W., BERTONE A., GÄRTNER J., MAYR E., MIKSCH S., SMUC M.: Hierarchical Temporal Patterns and Interactive Aggregated Views for Pixel-based Visualizations. In *Proc. of IV09* (2009), IEEE, pp. 44–49. 3

[Lam10] LAMMARSCH T.: *Facets of Time—Making the Most of Time's Structure in Interactive Visualization*. Ph.D. thesis, Vienna University of Technology, May 2010. Supervisors: Silvia Miksch (Vienna University of Technology), Daniel Keim (University of Konstanz). 3

[PC05] PIROLLI P., CARD S.: Sensemaking processes of intelligence analysts and possible leverage points as identified through cognitive task analysis. In *Proc. of the 2005 International Conference on Intelligence Analysis* (2005). 1

[SML*08] SMUC M., MAYR E., LAMMARSCH T., BERTONE A., AIGNER W., RISKU H., MIKSCH S.: Visualizations at First Sight: Do Insights Require Traininig? In *Proc. of USAB08* (2008), Springer, pp. 261–280. 1, 3

[SML*09] SMUC M., MAYR E., LAMMARSCH T., AIGNER W., MIKSCH S., GÄRTNER J.: To Score or Not to Score? Tripling Insights for Participatory Design. *IEEE Computer Graphics and Applications 29*, 3 (2009), 29–38. 1, 3

[TC05] THOMAS J., COOK K.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE, 2005. 1