KAVA-Time

Knowledge-Assisted Visual Analytics Methods for Time-Oriented Data

Wolfgang Aigner, Alexander Rind, and Markus Wagner

IC\M/T, Fachhochschule St. Pölten, Austria
wolfgang.aigner@fhstp.ac.at

Abstract. Visual analytics intertwines interactive visual interfaces with automated data analysis methods in order to support humans in data analysis. How visual analytics can leverage explicit knowledge from domain experts was investigated in the basic research project KAVA-Time. Within its scope, a theoretical model for integrating the users' knowledge into the visual analytics processes and two cases studies in the application domains IT security and clinical rehabilitation were developed.
Keywords: Visual analytics, explicit knowledge, malware analysis, physiotherapy.

1 Introduction

To keep pace with tremendously expanding malware families, IT-security experts have to explore large volumes of complex and heterogeneous data for subsequences of potentially malicious code. Even though the sheer quantity demands automated analysis methods, this process cannot be automated completely, since experts need to be in the loop to identify, correct, and disambiguate intermediate results [1, 2].

Many application fields besides malware analysis face this need for a coherent data analysis process intertwining human expertise and analytical processing, which knowledge-assisted visual analytics aims to address. The comparatively young field of "visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets" [3]. However, this endeavor is not straightforward as initial results from automated analysis are often trivial or irrelevant to the work of the domain experts. The expert user's knowledge from prior experience is an important asset that can be leveraged by both the user and the computer to improve the analytics process. While visual analytics environments are starting to include features to formalize, store and utilize such explicit knowledge, the mechanisms and degree to which these environments integrate explicit knowledge varies widely. Additionally, a theoretical model and formalization of this class of knowledge-assisted visual analytics environments is not available in the scientific community yet.

2 Scientific Results

Therefore, the basic research project KAVA-Time aims to close this gap by proposing a new theoretical high-level model of knowledge-assisted visual analytics (Fig. 1) [4]. The theoretical model builds upon van Wijk's operational model of visualization [5] and distinguishes between explicit knowledge in the machine space and tacit knowledge in the human space. It elaborates the processes involving knowledge [4]:

- Knowledge can be exploited to adjust the settings of visualization and intelligent data analysis. It can also be leveraged to provide guidance to the human.
- Tacit knowledge can be externalized, .i.e., made available to the machine through a dedicated knowledge specification interface or automatically inferred from the human's interaction with the machine.
- Explicit knowledge can be internalized by knowledge visualization or visualization of data that is simulated based on knowledge.
- Knowledge can be generated by the human cognition based on perception of visualized data or using automated data analysis.

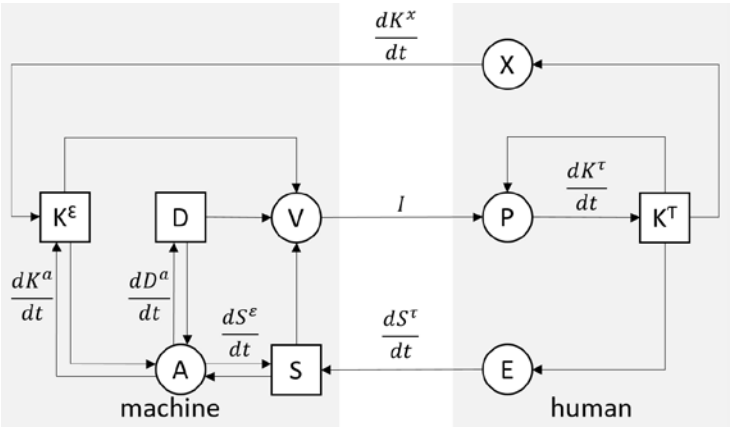

Fig. 1. Conceptual model of knowledge-assisted visual analytics [4]

Furthermore, KAVA-Time demonstrates the applicability of knowledge-assisted visual analytics in two fields involving time-oriented data: behavior-based malware analysis [6] and clinical gait analysis in ambulatory rehabilitation [7]. For each case study, the project team followed a user-centered design process to first characterize the analysis requirements, then develop a tailor-made software prototype (Fig. 2), and finally evaluate its usability and utility through IT-security professionals or physiotherapists respectively.

Feedback from domain experts collected in these evaluation studies underlined the capabilities of visual analytics in general. It also demonstrated the benefits of integrat-

ing explicit knowledge into the analysis process. In KAMAS, for example, malware analysts can externalize their knowledge by dragging call sequences into the knowledge store. Then, they can exploit explicit knowledge to filter or color-encode call sequences. Additionally, they can generate new tacit knowledge by identifying new characteristic sequences undisturbed from already known sequences.

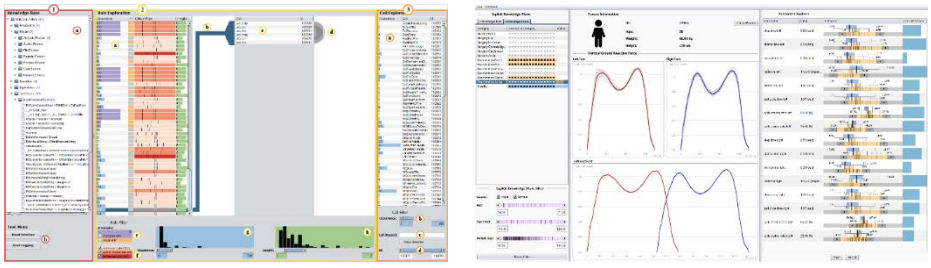

Fig. 2. VA prototypes to demonstrate the applicability: KAMAS prototype for behavior-based malware analysis [6] and KAVAGait prototype for clinical gait analysis [7].

3 Broader Effects

KAVA-Time has taken an active approach to transfer its results towards other scientists, industry, and the general public by addressing the different interests of these audiences:

- Besides open access publications, the research data prepared for and collected in the evaluation studies of KAMAS and KAVAGait were deposited in the university's long-term open access storage system.
- The cooperation partners involved in the user-centered design of both prototypes have showed interest to integrate knowledge-assisted visual analytics methods into their workflows. In the case of KAMAS, the necessary further development [8] started in a follow-up project with IKARUS Security Software.
- Visual analytics research was exhibited on three occasions at long-night-of-research events. Various visual analytics solutions including KAMAS and KAVAGait but also a baby name explorer [9] were showcased there.

Since its start in 2013, KAVA-Time has been the driver of innovation by establishing visual analytics as a new research area at FH St. Pölten and being the incubator for three large FWF- or FFG-funded projects (VALiD, VisOnFire, VAST). Today, visual analytics education is a part of several bachelor and master curricula. Wolfgang Aigner completed his habilitation at TU Wien and Markus Wagner his doctoral studies. Additional four colleagues are in the progress of their doctoral studies on visual analytics topics.

The visual analytics team at FH St. Pölten has established itself as recognized player in the international scientific community and serves in the program committees

of the top venues of the field (IEEE VAST, ACM CHI, EuroVis, EuroVA etc.). Recently, the team presented seven publications at IEEE VIS 2017, the most important scientific venue in visual analytics. Thus, KAVA-Time could significantly contribute to Austria's top ranking in the field of visual analytics.

Acknowledgements. KAVA-Time is supported by the Austrian Science Fund (FWF): P25489-N23. We thank our student researchers Andrea Haberson and Niklas Thür for their contribution to this project.

References

1. M. Wagner, W. Aigner, A. Rind, R. Luh, H. Dornhackl, K. Kadletz, and P. Tavolato. Problem characterization and abstraction for visual analytics in behavior-based malware pattern analysis. In K. Whitley, S. Engle, L. Harrison, F. Fischer, and N. Prigent, editors, Proc. 11th Workshop Visualization for Cyber Security (VizSec), pages 9–16. ACM, 2014. doi:10.1145/2671491.2671498.
2. M. Wagner, F. Fischer, R. Luh, A. Haberson, A. Rind, D. A. Keim, and W. Aigner. A survey of visualization systems for malware analysis. In R. Borgo, F. Ganovelli, and I. Viola, editors, Proc. Eurographics Conference on Visualization (EuroVis) – STARs, pages 105–125, 2015. doi:10.2312/eurovisstar.20151114.
3. D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. Mastering the Information Age – Solving Problems with Visual Analytics. Eurographics, Goslar, Germany, 2010.
4. P. Federico, M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In Proc. IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE, 2017. In press.
5. J. J. van Wijk. The value of visualization. In Proc. IEEE Visualization (VIS), pages 79–86, 2005. doi:10.1109/VISUAL.2005.1532781.
6. M. Wagner, A. Rind, N. Thür, and W. Aigner. A knowledge-assisted visual malware analysis system: Design, validation, and reflection of KAMAS. Computers & Security, 67: 1–15, June 2017. doi:10.1016/j.cose.2017.02.003.
7. M. Wagner, D. Slijepčević, B. Horsak, A. Rind, M. Zeppelzauer, and W. Aigner. KAVAGait: Knowledge-Assisted Visual Analytics for Clinical Gait Analysis. IEEE Transactions on Visualization and Computer Graphics, 2018. In press. Preprint published as arXiv:1707.06105.
8. N. Thür, M. Wagner, J. Schick, C. Niederer, J. Eckel, R. Luh, and W. Aigner. BiG2-KAMAS: Supporting knowledge-assisted malware analysis with bi-gram based valuation. In Poster Proc, 14th Workshop on Visualization for Cyber Security (VizSec), Phoenix, AZ, USA, 2017.
9. M. Wagner, K. Blumenstein, A. Rind, M. Seidl, G. Schmiedl, T. Lammarsch, and W. Aigner. Native cross-platform visualization: A proof of concept based on the Unity3D game engine. In Proc. Int. Conf. Information Visualisation (IV16), pages 39–44. IEEE, 2016. doi:10.1109/IV.2016.35