Interactive visual event-sequence mining

CENIIT project final report

Katerina Vrotsou <katerina.vrotsou@liu.se>

Project background and industrial motives

The CENIIT project "Interactive visual event-sequence mining" is concerned with one of the main drawbacks of most data mining algorithms in general and sequence mining algorithms in particular. Namely, the fact that most algorithms operate as a "black box" and the contribution of the end user, that aims to benefit from their results, is limited to adjusting some initial parameters which act as constraints on them. Long lists of patterns are then output from the computation of these algorithms many of which are irrelevant to the end user who needs to scan and assess them.

The core issue we address within this project is to develop approaches to sequential pattern mining that deeply embed interaction and domain knowledge within the mining algorithm to facilitate exploratory sequence mining. In this way, interestingness of sequence mining results is enhanced. Interestingness in this context implies the relevance of the results to the analysis questions of the domain expert performing the analysis. To accomplish this we have been investigating and developing visual analytics approaches that combine modern interactive visualization techniques with algorithmic computation in order to create a "transparent box" execution model of analysis.

The research undertaken in this project is motivated by the vast number of systems which produce high-dimensional event-sequence data, many of them being significant to both society and industry. Examples can be found in industrial processes where monitoring and early warning systems produce sequences of alarm-events; in eye-tracking experiments as sequences fixations on areas of interest; in the medical sector through the use of medical records for diagnostics and treatment planning among many others.

Summary of most important scientific results

This project has over its course been evolving in two paths. One path has been focusing on core research issues and has been producing results in terms of development of new algorithms and interactive analysis methods that are applicable to a wide range of areas as well as theoretical contributions (publications [1-3],[5-7],[10-14],[17-18],[22-29]). A second part has been concerned with the use of the researched methods for performing analyses and identifying new directions for continued research (publications [4], [8-9],[15-16], [19-21],[30]). Through this process, several extensions into new application areas have emerged and new related projects have commenced.

A core scientific contribution produced during this project has been the exploratory sequence mining approach presented in [17,28] and the accompanying prototype system, ELOQUENCE, implementing the approach. In continued research, the approach was extended to handle long duration sequences [10] and the identification of patterns within and between sequences. In addition, the definition of similarity between event-sequences and their comparison based on this has been explored [27]. These have provided a solid

base for our continued research on interactive visual event-sequence analysis and have been applied and extended in a number of application oriented projects and problems. Particularly, through our collaboration with the Swedish Air Navigation Provider (LFV), a central research focus has been put on the analysis of eye-tracking data and the identification and analysis of patterns of visual scanning behaviour where we have made and reported several contributions, building on combinations of computational and visualization approaches, over the past years [1,3,7,11,13,14,18,23,24]. Through this research additional challenges have been identified concerned with the effective analysis of long and noisy event sequences. To address these challenges, two novel approaches have been proposed recently; the first is concerned with the interactive context-aware simplification of event sequences to enable the progressive improvement of mining results [1] and the second is concerned the identification of overarching behavioural phases as patterns from the event sequences [29].

Two additional theoretical contributions made in the frames of this project have been concerned with the potential of sequence mining for untangling complexity in complex control processes, as that of air-traffic control [6], and the creation of a taxonomy of levels of automation for visual analytics with examples in event-sequence analysis systems [2].

Complementing the above outlined core research several collaborative papers have been produced applying aspects of interactive visual event sequence analysis to a number of diverging areas [4,5,8,9,15,16,19,20,21,25].

Summary of degrees and promotions

Prithiviraj Muthumanickam received his PhD with his doctoral thesis entitled “Data Abstraction and Pattern Identification in Time-series Data” in 2019. Prithiviraj, who was co-supervised by Katerina Vrotsou and Aida Nordman, was associated to the project but was funded by a Swedish research council grant (“A user-centered information visualization framework for detecting and exploring change in high dimensional temporal data”).

Katerina Vrotsou was promoted to Senior lecturer in Information Visualization (Universitetslektor) in April 2020 and to Docent in Visualization and Media Technology in January 2021.

Masters thesis performed within the project

There have been two masters thesis that have been directly related to the project: (1) “Visualization and analysis of diagnostic/maintenance data for promoting lifetime management of gas turbines” conducted by Jonas Petersson in 2015, and (2) “Interactive Visualization System Supporting User Guided Constraint-based Sequence Mining” conducted by Joakim Deborg in 2016.

Persons funded by the project

The project over its course has been mainly funding Katerina Vrotsou (PI), Aida Nordman (co-applicant) and occasionally a research engineer.

Industrial collaboration

- Main industrial collaborator of the project since 2018 has been the Swedish Air Navigation Service Provider (ANSP) 'Luftfartsverket' (LFV). During this time a number of parallel projects have emerged:

  [1] *RESKILL* project, in collaboration with LFV and Swedish Maritime Administration, is concerned with the identification and exploration of visual scan patterns of air-traffic controllers in tower control and maritime pilots with a focus in training.

  [2] *OVaK* ('Objektiv verifiering av kompetens mha avancerad automation') is concerned with exploring interactive sequence mining for identifying scan patterns of en-route ATCOs in air-traffic control centres with a focus on how ATCOs handle conflicts in air-space.

  [3] "Interactive visual data analysis of system monitoring data in Air Traffic Management" (short name: *VDM*) is concerned with researching the use of interactive visual sequence mining techniques applied to system monitoring data and related metadata collected from various air traffic management systems that LFV has across the country.

  In most of our work with LFV, a primary focus has been in creating tools that can be used in training and competence assurance of air traffic controllers.

- Swedish Maritime Administration through the RESKILL project (see point [1] above).

- Siemens Industrial Turbomachinery AB was concerned with the exploration and identification of patterns in complicated operation and maintenance history data of gas turbines.

New research group and research development

The funding provided by CENIIT has been decisive for the advancement and development of the main applicant, Katerina Vrotsou. The grant was received short after she finished her post-doc and has contributed to her development and vision as a researcher. It has allowed her to devote time to this research area and what started as a small niched project has resulted in follow-up funding for continuing and extending the research in new directions. The She has led two related projects funded by Trafikverket; OVaK (2020-2022), VDM (2020-2021). Also, in 2020 she was awarded a starting grant from the Swedish Research Council (VR) for the project "Progressive Interactive Event Sequence Analytics: Combining Explorative Mining and Visual Reasoning" to continue the research started by the CENIIT project. The group currently working actively with Katerina Vrotsou consists of a PhD student, a post-doc and a senior lecturer. In parallel, she has had the opportunity to broaden her research vision and expand the interactive sequence mining ideas of the CENIIT project to other fields and aspects of interactive visual data analysis, including visual text analytics and immersive visual analytics.

Publications list

[1] P. Yu, A. Nordman, L. Meyer, S. Boonsong, K. Vrotsou. 'Interactive Transformations and Visual Assessment of Noisy Event Sequences: An Application in En-Route Air Traffic Control'. To appear in 16th IEEE Pacific Visualization Symposium (PacificVis), 2023.

[2] V. Domova, K. Vrotsou. 'A Model for Types and Levels of Automation in Visual Analytics: a Survey, a Taxonomy, and Examples', IEEE Transactions on Visualization and Computer Graphics, 2022.

[3] L. Meyer, S. Boonsong B. Josefsson, A. Nordman, K. Vrotsou, C. Westin, K. J. Klang and J. Lundberg. 'Mapping the Decision-Making Process of Conflict Detection and Resolution in En-Route Control: An Eye-tracking based approach'. In *41st Digital Avionics Systems Conference (DASC)*, 2022.

[4] K. Orban, K. Vrotsou, K. Ellegård, and L.-K. Erlandsson. 'Assessing the use of a portable time-geographic diary for detecting patterns of daily occupations'. *Scandinavian Journal of Occupational Therapy*, 29(4): 293-304, 2022.

[5] V. Ceccato, D. J. Wiebe, K. Vrotsou, U. Nyberg, and A. Grundberg. 'The situational conditions of suicide in transit environments: An analysis using CCTV footage'. *Journal of Transport Health*, 20:100976, 2021.

[6] M. Nylin, K. Vrotsou, M. Bång, and J. Lundberg. 'Untangling complexity in real-time control processes interactions'. In *Emergent Interaction Workshop of the 2021 ACM CHI Virtual Conference on Human Factors in Computing Systems*, 2021.

[7] L. Meyer, B. Josefsson, K. Vrotsou, C. Westin, and J. Lundberg. 'Evaluation of an AOI mapping and analysis tool for the identification of visual scan pattern'. In *40th Digital Avionics Systems Conference (DASC)*, 2021.

[8] G. Andersson, K. Vrotsou, A. Denhov, A. Topor, P. Bulow, and K. Ellegård. 'A diversity of patterns: 10-year trajectories of men and women diagnosed with psychosis for the first time. A time-geographic approach'. *Moravian Geographical Reports*, 28(4):283–298, 2020.

[9] K. Vrotsou and W. Glad. 'Visualizing thermal comfort in residential passive house designs'. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, e-Energy '21, page 412–416, New York, NY, USA, 2021. Association for Computing Machinery.

[10] K. Vrotsou, A. Nordman. 'A window-based approach for mining long duration event-sequences'. In *EuroVis Workshop on Visual Analytics (EuroVA)*, Norrköping, Sweden, May 2020.

[11] P. Muthumanickam, A. Nordman, J. Helske, J. Johansson, and M. Cooper. 'Comparison of attention behaviour across user sets through automatic identification of common areas of interest'. In *53rd Hawaii International Conference on System Sciences*, Maui, Hawaii, January 2020.

[12] P. Muthumanickam. 'Data Abstraction and Pattern Identification in Time-series Data'. Vol. 2030. Linköping Studies in Science and Technology, Dissertation No. 2030, Linköping University Electronic Press, 2019.

[13] C. A. L. Westin, K. Vrotsou, A. Nordman, J. Lundberg, and L. Meyer, 'Visual scan patterns in tower control: Foundations for an instructor support tool'. SESAR Innovation Days, December 2019.

[14] K. Muthumanickam, A. Nordman, L. Meyer, S. Boonsong, J. Lundberg, and M. Cooper. Analysis of long duration eye-tracking experiments in a remote tower environment. In *13th USA/Europe air traffic management R&D seminar*, Vienna, Austria, 2019.

[15] G. Andersson, K. Ellegård, P. Bülow, A. Denhov, K. Vrotsou, C.-G. Stefansson, and A. Topor, 'A longitudinal study of men and women diagnosed with psychosis: trajectories revealing interventions in a time-geographic framework', *GeoJournal*, 2019.

[16] D. Anaby, K. Vrotsou, U. Kroksmark, and K. Ellegård, 'Changes in participation patterns of youth with physical disabilities following the Pathways and Resources for Engagement and Participation intervention: A time-geography approach', Scandinavian Journal of Occupational Therapy, 2019.

[17] K. Vrotsou and A. Nordman, 'Exploratory visual sequence mining based on pattern-growth', IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 8, pp. 2597–2610, 2019.

[18] P. Muthumanickam, K. Vrotsou, A. Vitoria, J. Johansson, and M. Cooper, 'Identification of Temporally Varying Areas of Interest in Long-Duration Eye-Tracking Data Sets', *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 87–97, 2018.

[19] M. Haga, K. Vrotsou, and E. Bredland, 'Visualizing Physical Activity Patterns among Community-Dwelling Older Adults: A Pilot Study', *Sports*, vol. 6, no. 4, 2018.

[20] K. Vrotsou et al., 'A time-geographic approach for visualizing the paths of intervention for persons with severe mental illness', *Geografiska Annaler. Series B, Human Geography*, vol. 99, no. 4, pp. 341–359, 2017.

[21] V. Ceccato, D.J. Wiebe, B. Eshraghi, K. Vrotsou. Women's Mobility and the Situational Conditions of Rape: Cases Reported to Hospitals. *Journal of Interpersonal Violence*, 2017.

[22] P. K. Muthumanickam, K. Vrotsou, M. Cooper, and J. Johansson. Shape grammar extraction for efficient query-by-sketch pattern matching in long time series. In *IEEE Conference on Visual Analytics Science and Technology*, 2016.

[23] P. K. Muthumanickam, Å. Svensson, A. Nordman and K. Vrotsou. Analyzing behaviour of air traffic controllers: An exploratory visual sequence mining approach. Poster in the *Sixth SESAR Innovation Days*, 2016.

[24] P.K. Muthumanickam, C. Forsell, K. Vrotsou, J. Johansson, M. Cooper. Supporting Exploration of Eye Tracking Data: Identifying Changing Behaviour Over Long Durations. Proceedings of *Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 70-77, 2016.

[25] G. Mattioli, J. Anable, K. Vrotsou. Car dependent practices: Findings from a sequence pattern mining study of UK time use dat. *Transportation Research Part A: Policy and Practice*, 89, 56-72, 2016.

[26] P. K. Muthumanickam, K. Vrotsou, M. Cooper, and J. Johansson. Smart series: Sketch based matching through approximated ratios in time series. In *IEEE Conference on Visual Analytics Science and Technology*, 2015.

[27] K. Vrotsou, A. Ynnerman, M. Cooper. Are we what we do? Exploring group behaviour through user-defined event-sequence similarity. *Information Visualization* 13 (3), 232-247, 2014.

[28] K. Vrotsou and A. Nordman. Interactive visual sequence mining based on pattern-growth. In *IEEE Conference on Visual Analytics Science and Technology*, 2014.

Submissions currently prepared or under submission

[29] A. Nordman, K. Vrotsou K. J. Klang, J. Lundberg and L. Meyer. 'Extraction of CDR work phases from eye-tracking and simulator logs: a topic modelling approach'. Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023), 2023.

[30] C. Westin, A. Nordman, K. Vrotsou, G. Schnücker, L. Meyer, A. Johannesson. 'What is the pattern? Visual scan patterns in vessel traffic services'.