Exploring Lekagul Sensor Events using Rules, Aggregations, and Selections

Bram C.M. Cappers

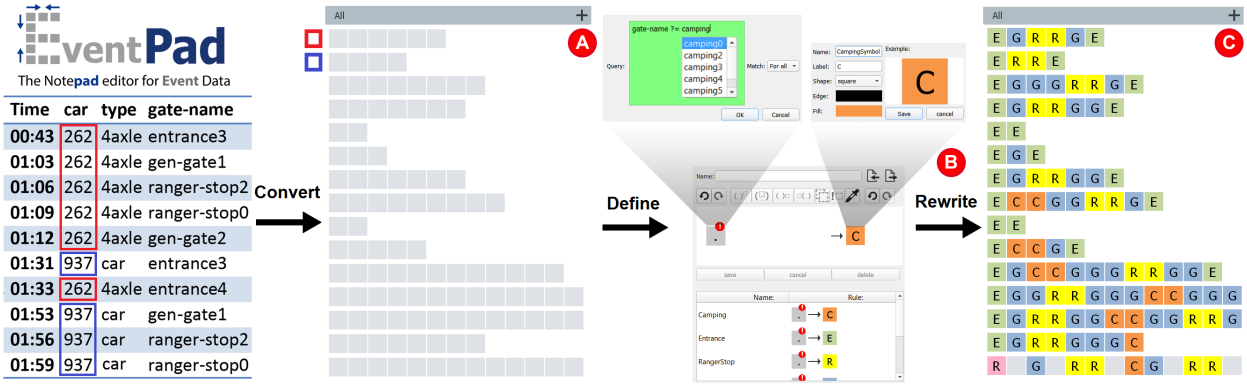

Fig. 1. The Eventpad dataflow model for multivariate event sequence exploration: Table records are grouped by an attribute of choice and represented as series of glyphs (A). Users can visually encode sequences according to attributes of interest by constructing one or more rules. Rules are constructed using multivariate regular expressions. Users can design their own glyphs to highlight points of interest. C) Event collection after rule rewriting.

Abstract— In this paper we demonstrate how we can study multivariate event sequences in the VAST Mini Challenge 1 data set using our system Eventpad, a notepad editor for event data. We illustrate the effectiveness of multivariate regular expressions, pattern aggregations, and selections to define custom events of interest, discover patterns within sequences, and study differences between sequences. Finally, we discuss our analysis process and summarize some patterns and anomalies we discovered in the data set.

Index Terms—Event Visualization, Multivariate Events, Regular Expressions, Sequence Alignment, Interaction

1 Introduction

The Lekagul Natural Preserve records sensor events for every vehicle driving through gates inside the habitat. Besides a timestamp and car-id these events store additional multivariate data such the type of vehicle it corresponds to and the gate name it passed. In order to study normal and anomalous behavior in these multivariate event sequences, we need to become aware of patterns within the sequences, between sequences and inside event properties. To achieve this we developed Eventpad [2]: a notepad editor for event data.

Eventpad is a novel system designed to simplify and study patterns in multivariate event sequences. Similar to notepad editors, the system uses find and replace functionality to discover patterns inside the data. To study commonalities and differences between such sequences, the system enables users to simultaneously explore sequential patterns alongside multivariate data. In order to achieve this, the system relies on three concepts, namely

- *rules* to highlight, find, and compress event sequences. For this we extended regular expressions to support multivariate data,
- *aggregations* to discover similarities between sequentially similar but structurally different sequences through clustering, partitioning, sorting, and alignment, and
- *selections*, to study differences in multivariate data using selections of interest.

- *Bram C.M. Cappers is with Eindhoven University of Technology. E-mail: b.c.m.cappers@tue.nl*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

To facilitate interplay between these concepts, the Eventpad system consists of five views, namely:

- a Sequence view (Figure 3B-1), representing every event sequence a series of glyphs,
- an Alignment view (Figure 2B), aligning event sequences of interest using Multiple Sequence Alignment,
- an Attribute view (Figure 3B-2), to study patterns inside multivariate data of selections of interest using scented widgets,
- a Rule view (Figure 3B-3), showing the impact and ordering of applied rules in the visualization, and
- a Context view [1] (Figure 3B-4), to store selections of interest for further investigation.

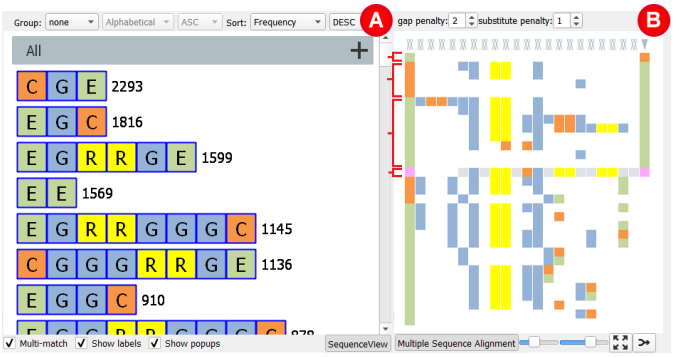

Fig. 2. A) Most frequent daily patterns clustered by their visual representation. B) Applying Multiple Sequence Alignment to event sequences of interest enables user to discover similarities between travel patterns.

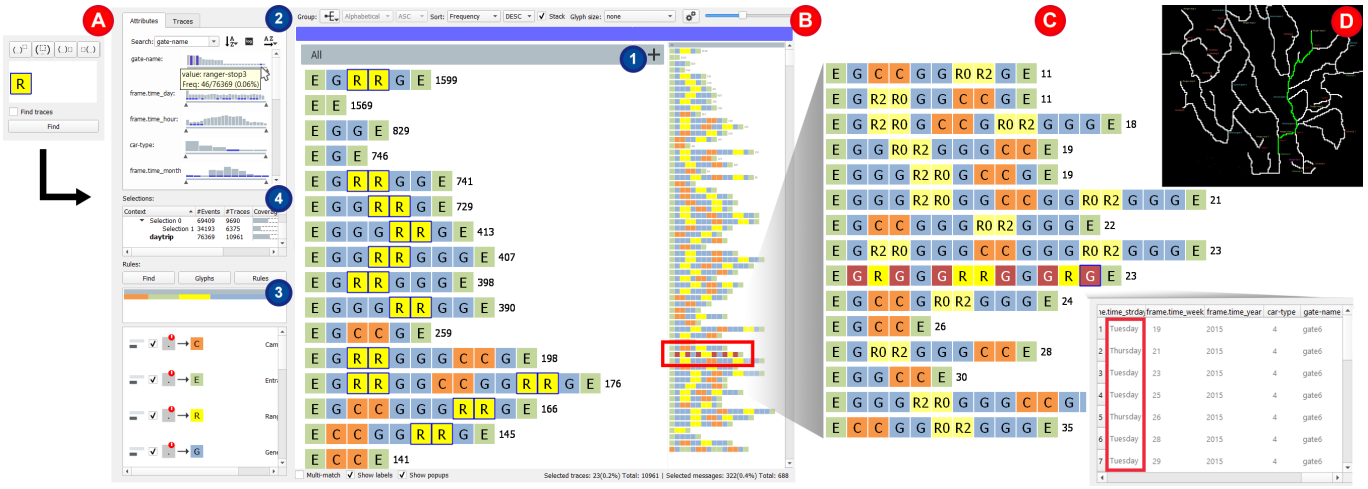

Fig. 3. A) Searching for rangerstops in sequences. B) Graphical user interface of the implemented prototype: 1) Sequence view represents sequences as series of glyphs. 2) The attribute view shows per-attribute trends and patterns in selections of interest. 3) The rule view shows the coverage and ordering of applied rules. 4) the context view stores selections of interest throughout exploration. C) Discovery of 4axle trucks driving through illegal gates. D) Roadmap of Lekagul.

2 Exploration

We initially start the search for daily patterns by grouping the data by car-id per day. To gain insight in travel patterns, we construct five rules where camping events are colored in orange, entrances in green, general-gates in blue, rangerstops in yellow, and rangerbase events in pink.

For the inspection of frequent daily patterns in the data, we cluster sequences based on their visual representation and sort them by frequency (Figure 2B). Applying Multiple Sequence Alignment on the most frequents sequences enables us to identity four main patterns, namely vehicles entering, leaving, and driving through the preserve along with ranger traffic (Figure 2B).

To study enter and leave behavior of vehicle types, we group the traffic by car-id only. We construct a rule that compresses all enter and exit patterns of vehicles in the data into purple glyphs (Figure 4A). We do this by stating that in between two entrance events, no other entrance events are allowed (Figure 4B). Inspecting only these patterns shows that certain vehicles visit the preserve multiple times in a year (Figure 4A). This also shows that ranger vehicles never leave the preserve. Selecting the long sequence in Figure 4A and disabling the constructed rule shows that this sequence corresponds to a 2axle truck driving systematically between entrance 4 and camping 4 during high season only (Figure 4C).

According to the challenge description, only ranger vehicles are allowed to travel through rangerstops. In Eventpad we can easily verify this statement by searching for sequences with rangerstops whose car type differs from ranger vehicles (Figure 3A). This reveals 23 cases where 4axle trucks are driving midnight between rangerstop 3 and entrance 3 (Figure 3C). Inspection of the multivariate data in these sequences in a tabular view shows that these vehicles only drive on Tuesdays and Thursdays. For a demonstration of the system in practice, we refer to the supplementary video [1].

3 Conclusion

We have shown the effectiveness of Eventpad to quickly gain insight in the VAST 2017 Mini Challenge 1 data set. The ability to visually encode event properties in sequences using rules enables users to quickly discover patterns inside sequences. Pattern aggregations and selections enable users to study commonalities and differences between sequences while staying aware of high-level phenomena in the data set.

Using rules, aggregations, and selections, we discovered that vehicles on certain roads drive too fast and enter locations in the middle of the night for which they are not authorized. In addition, the presence of systematic travel activity across the entire preserve during high-season can also disturb the wildlife in Lekagul.

[1] https://www.youtube.com/watch?v=IBgJ3R9cAvQ

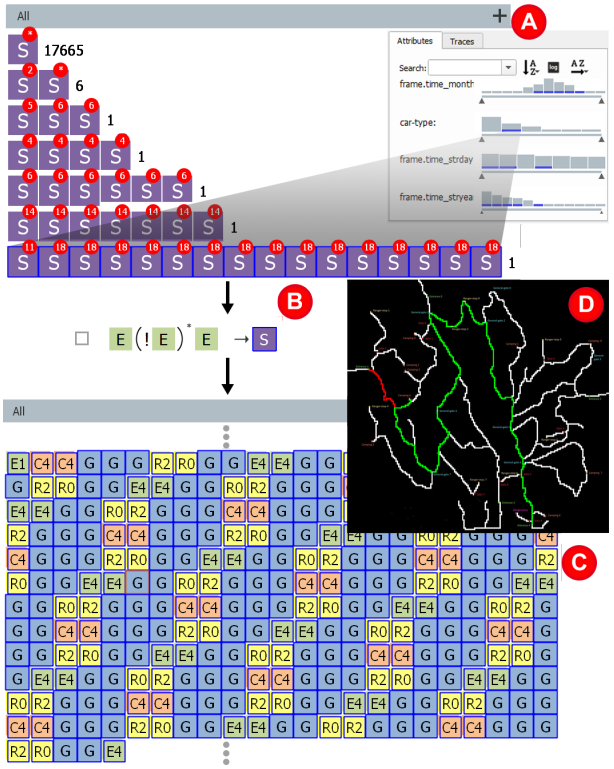

Fig. 4. A) Number of times vehicles enter and leave the preserve. B) Compression rule for the discovery of visitor patterns (disabled). C) Systematic traveling of a 2axle truck during high season (D).

Acknowledgments

This work is funded by SpySpot, a project in the Cyber Security program of Netherlands Organisation for Scientific Research (NWO).

References

[1] B. C. M. Cappers and J. J. van Wijk. Understanding the Context of Network Traffic Alerts. In *Visualization for Cyber Security (VizSec), 2016 IEEE Symposium on*, pp. 1–8. IEEE, 2016.

[2] B. C. M. Cappers and J. J. van Wijk. Exploring Multivariate Event Sequences using Rules, Aggregations, and Selections. *To be published in IEEE Transactions on Visualization and Computer Graphics*, 2017.