Identifying candidate causal relationships in fish movement patterns

Susanne Bleisch[1], Matt Duckham[1], Patrick Laube[2], Jarod Lyon[3]

[1]University of Melbourne, Parkville 3010, Victoria, Australia
Tel. +61 3 8344 6935 Fax +61 3 8344 4616
matt@duckham.org, http://ambientspatial.net/amsi
[2]University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[3]Arthur Rylah Institute, 123 Brown St, Heidelberg, Victoria, Australia

KEYWORDS: causation, events, processes, context aware movement analysis,

1. Introduction

Although there has been considerable research on the analysis of the *second order effects* of movement (i.e., the analysis of the patterns, similarities, or characteristics of trajectories), Gschwend and Laube (2012) argue that not enough research has addressed the analysis of *first-order effects* (i.e., the impact of the environmental context in which movement takes place). In this paper we outline ongoing research investigating an important and challenging aspect of context-aware movement analysis: causation. Specifically, the research aims to explore techniques for mining dense spatiotemporal data about movement patterns and environmental changes to assist domain experts in identifying and testing hypotheses about possible causal relations between movement events and environmental events.

1.1 Motivation: Fish movement

A major environmental monitoring project in the Murray River, south eastern Australia, has established over the past five years a network of 18 logging towers, capable of tracking the movement of more than 1000 tagged fish between different zones in the river (Koehn et al. 2010). The project is providing important insights into the effects of environmental management of the river to improve fish habitats. However, this rich data source may contain many further important insights about the ecology of this environmentally and economically vital river system. For example, do moon cycles, high river flows, or rain events *cause* certain patterns of long-range movement in tagged fish? Or do low temperatures lead to fish staying in home zones of the river? In general, our contention is that the (first-order) *context* of animal movement, and in particular the causal context, is typically more important to ecological domain experts than the (second-order) characteristics of the movement itself.

2. Causation

Causal relationships are relatively infrequently studied in the context of geographic information science (but see, e.g., Allen et al 1995, El-Geresy et al. 2002 amongst others), no doubt in part due to the difficulty in arriving at a consensus on the definitions of causal relationships, even in philosophy. We adopt for the ontological foundations of this work the approach of Galton (2012), summarized in Figure 1. In starting from solid ontological foundations, our hope is to make our approach flexible enough to be useful in a range of applications beyond our specific example of fish movement.

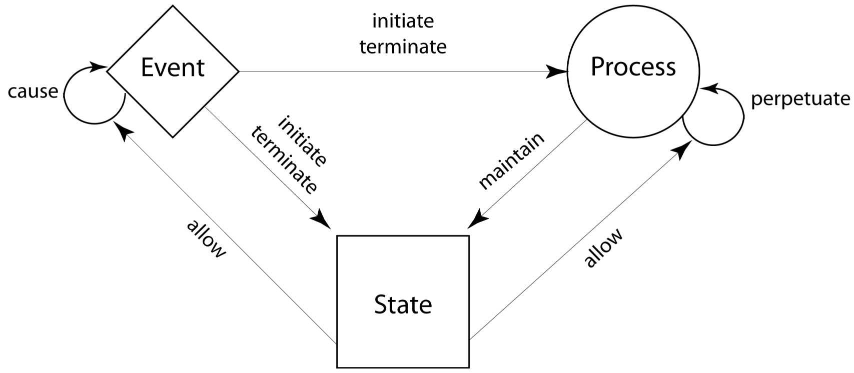

Figure 1. Causal-like relationships amongst states, events, and processes, after Galton 2012

In short, after Galton (2012):

- Only events (defined as temporally bounded "happenings" where one or more participants in that event change) may strictly *cause* other events.
- Events may *initiate* or *terminate* states (causal-like relationships) and may *initiate* or *terminate* processes (similar to events, but an open-ended activity, like "movement").
- Processes can *perpetuate* other processes (analogous to an event causing and event)
- States (of the world) only affect causation in as much as they can *allow* events to cause other events or processes to perpetuate other processes.

Based on this ontology, our approach further aim to identify event sequences where consecutive events *share some spatiotemporal part*. The intuition here is that in order for one spatial event to cause another, they must share at least one point in space and time.

2.1 Example

In our motivating example of fish movement, events, processes, and states and their interrelationships can all play important roles in context aware movement analysis. For example, the *state* of high flows in parts of the river will be initiated by an *event*: the start of a high river flow (i.e., the transition from not high to high flow). A subsequent full-moon *event* may *cause* the start of a fish migration *event*, this causation being *allowed* by the *state* of high river flow. Spatiotemporal co-location is essential to this causal relationship. Fish in either a. other parts of the river, which are not in a state of high flow; or b. that river zone but at another time when it is not in a state of high flow, may consequently not be allowed to experience the start of a fish migration event. The start of the fish migration event *initiates* a *process* of fish migration, with fish swimming long distances upstream to spawn. Figure 2 summarizes these relationships diagrammatically.

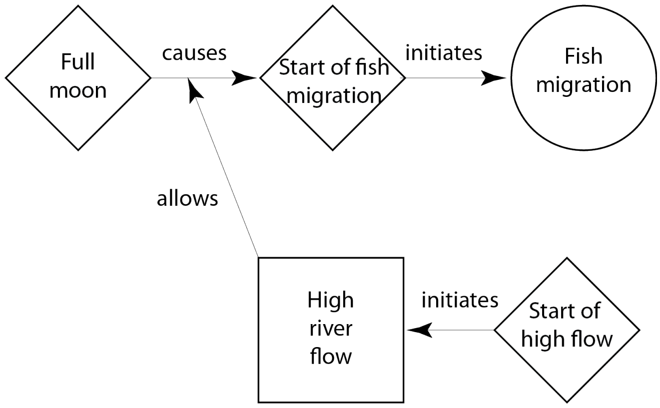

Figure 2. Example ontology of fish migration causation (cf. Figure 1)

Clearly, it would be beyond any of today's data mining or artificial intelligence techniques to derive such causal *explanations* given only data about fish movements and their environmental context. Nevertheless, our experience is that domain experts typically already possess many (competing) hypotheses about what causal relationships may explain patterns in a data set (even if they are unlikely to be phrased in the somewhat artificial language of ontology used above). Our aim in this research is to provide tools that can assist domain experts in exploring dense data sets concerning fish movement and associated environmental factors.

3. Sequence mining

Figure 3 illustrates our approach to identifying potential causal relationships in data from our motivating example of fish movements in the Murray River (section 1.1). In summary:

- The data generated from the fish-tracking project consists of tuples of the form (*t*, *l*, *o*) where *t* is a timestamp; *l* is a location (in the form of a river zone); *o* is an observation of a fish's identifier in that zone. Environmental data confirms to the same structure, except that the observations are of environmental parameters, including temperature, river flows, moon phase, and so forth. In both cases, observations are per-day over a four-year period.

- The raw data is categorized into sequences of atomic events experienced by each fish over time, such as the movement from one zone to another, or the start of a high temperature event in its zone. The formulation of event categories requires human domain expertise (e.g., a high temperature event occurs when water temperature rises by more than 1 degree above seasonal average).

- At this point, the user may choose to explore the data, using an existing data mining technique to mine frequent sequences of events from the data. Sequence mining is closely related to association rule mining. In our approach, we reuse existing sequence mining techniques and software, based on the TraMineR sequence mining package for the R statistics language (Gabadinho et al, 2011). The approach generates an exhaustive set of frequent event sequences (candidate causal relationships), which can be ranked by support (identified sequence as a proportion of all sequences).

- Whether through exploration of the mined frequent sequences, or directly through pre-existing expertise, it is to be expected that users may derive higher level hypotheses about causal relationships between events and processes. TraMineR also supports sequence mining of user-defined sequences, of atomic or derived aggregate events. An example of derived aggregate event would be a long-range movement, such as a sequence of four upstream zonal transitions within some time period. Further statistical analysis may be applied to determine whether particular identified sequences may have occurred by chance (hypothesis tests).

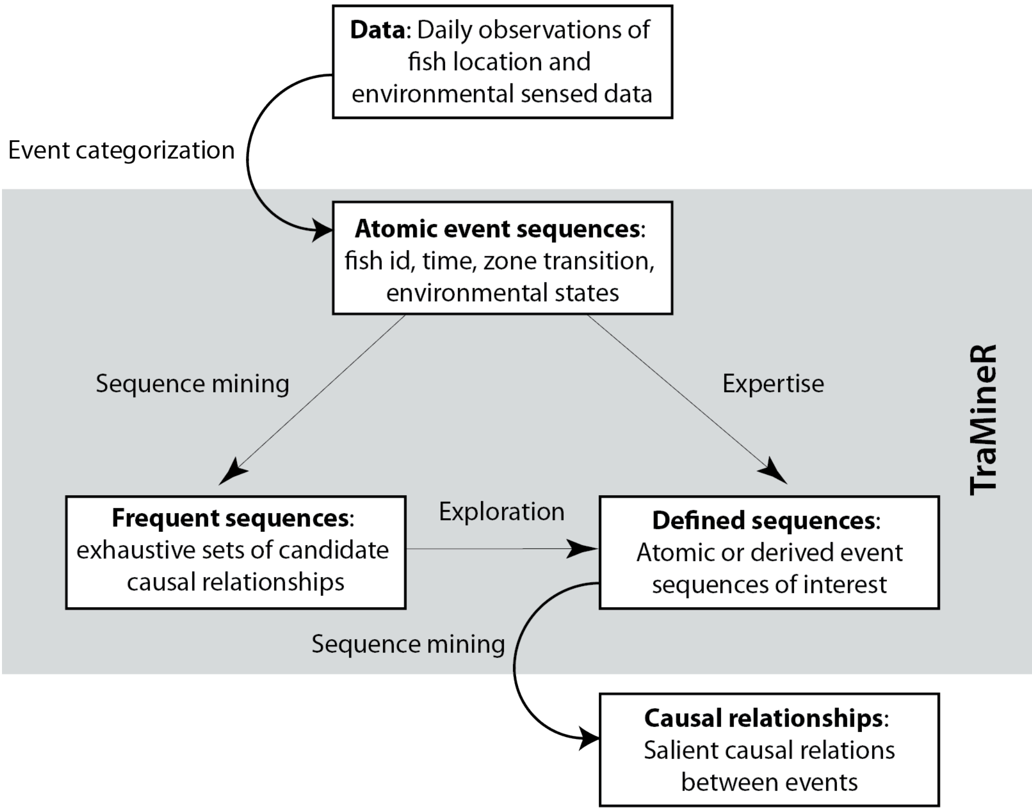

Figure 3. Summary of context-aware fish movement analysis process

4. Preliminary results and outlook

Refinement, standardization, and experimental evaluation of our approach is currently ongoing. However, early results are encouraging. Example of some of the statistically significant candidate causal relationships identified from the data include:

- Full and new moons are followed by downstream fish movements with a lag of up to two days (i.e., there is no significant increase in fish movements immediately on full or new moon days).
- High rain events tend to be followed by (upstream or downstream) fish movement, with a similar lag of up to two-days.
- Long range downstream or upstream movement tends to follow high (but not low) temperature events.

In addition to thorough experimental analysis of candidate causal events identified by the approach, current short-term research goals include:

- Further precise formal and ontological definition of the entities and relationships that can be identified using this process;
- The development of a user interface to efficiently construct event categorizations and interact with mined sequences; and
- Extensions to other closely related applications, including marine tracking data about shark movements and transportation data about vehicle movements.

5. Acknowledgements

This research is supported by funding from the Australian Research Council (ARC) under the Discovery Projects Scheme, project DP120100072 “From environmental monitoring to management: Extracting knowledge about environmental events from sensor data.”

References

Allen E, Edwards G and Bedard Y (1995) Qualitative Causal Modeling in Temporal GIS. In *Proceedings COSIT Conference*, Springer, Berlin, pp 397–412.

El-Geresy B, Abdelmoty AI and Jones CB (2002) Spatio-Temporal Geographic Information Systems: A Causal Perspective. In Manolopoulos and Pavol (Eds.) *Proceedings of the 6th East European Conference on Advances in Databases and Information Systems* (ADBIS '02), Springer, Berlin, pp 191–203.

Gabadinho A, Ritschard G, Müller NS, and Studer M (2011) Analyzing and visualizing state sequences in R with TraMineR, *Journal of Statistical Software*. **40**(4), 1–37.

Galton A (2012) States, Processes and Events, and the Ontology of Causal Relations In Donnelly and Guizzardi (Eds.) *Formal Ontology in Information Systems*, IOS Press, Amsterdam, pp 279–292.

Gschwend C and Laube P (2012) *Challenges of Context-Aware Movement Analysis – Lessons learned about Crucial Data Requirements and Pre-processing*. Proceedings of GISRUK 2012.

Koehn J, Nicol S, McKenzie J, Lieschke J, Lyon J and Pomorin K (2008) Spatial ecology of an endangered native Australian Percichthyid fish, the trout cod Maccullochella macquariensis. *Endangered Species Research*, **4**, 219–225.

Biography

Susanne Bleisch is a postdoctoral research fellow at the Department of Infrastructure Engineering, University of Melbourne, Australia. Her research focuses on using geovisualization approaches to better understand sensor network data.

Matt Duckham is an Associate Professor and ARC Future Fellow at the Department of Infrastructure Engineering, University of Melbourne, Australia. His research focuses on distributed and robust computation with uncertain spatial and spatiotemporal information.

Jarod Lyon manages the Restoration Ecology Research Program at the Arthur Rylah Institute, Department of Sustainability and Environment, Victorian Government, Australia. His work focuses on increasing knowledge of river rehabilitation techniques and increasing numbers and range of key threatened fish species throughout their natural range using best available restoration science.

Patrick Laube is a Lecturer at the Department of Geography, University of Zurich, Switzerland. Patrick’s research focuses on the processing of spatiotemporal information with the aim of better understanding of the dynamics in natural and built environments