Identifying perpetuation in processes driving fish movement

Matt Duckham*[1], Antony Galton†[2] and Alan Both‡[1]

[1]University of Melbourne, Australia
[2]University of Exeter, UK

January 9, 2015

Summary

This extended abstract explores ongoing work that is developing new models and algorithms capable of identifying the environmental drivers of human and animal movement. Specifically, the paper presents an algorithm able to identify the perpetuating conditions for movement: those ranges of environmental variables that are necessary for movement to occur. Our algorithm is tested on fish movement data from a large, long-term ecology study in Australia, combined with environmental data about water temperatures, levels, and salinity. The results demonstrate the types of rules that can generated from real movement patterns using our algorithm.

KEYWORDS: causation, processes, perpetuation, context-aware movement analysis, environmental monitoring, data mining.

1 Introduction

Context-aware movement analysis (CAMA) aims to relate movement to the underlying geographic context in which that movement is embedded (Laube, 2014). Understanding how geographic space drives movement patterns is arguably of much greater importance in most applications than analyzing the geometry of the movement patterns (a perspective often ignored by traditional movement analysis approaches, cf. Laube et al., 2005; Buchin et al., 2011). In this extended abstract we explore ongoing work developing techniques capable of identifying the environmental drivers of movement. Specifically, the work reported concerns the identification of perpetuating conditions for movement: those ranges of environmental conditions that are necessary for movement to occur, even though these conditions may not directly be the causes of movement. The approach, tested on fish movement data from a large long-term ecology study in Australia, demonstrates the types of rules generated from real movement patterns.

*mduckham@unimelb.edu.au
†apgalton@ex.ac.uk
‡alan.both@gmail.com

2 Perpetuation

Our algorithm for identifying perpetuation is a component of a larger attempt to identify other types of causation too, including previous work on identification of causal events in fish movements (Bleisch et al., 2013, 2014). Our general model of causation has as its primary focus the relationships between events, some of which are *causes* while others are *effects*. The most general form of causal rules we are working towards handling is:

$$[\mathsf{Causes} \mid \mathsf{Conditions}] \Rightarrow \mathsf{effect\ after\ Delay},$$

where

- $\mathsf{Causes} \subset \mathcal{C}$, $\mathcal{C}$ is a set of causes (environmental events),
- $\mathsf{Conditions}$ is a set of *conditions*, where each condition is a triple $c = (p_c, v_c^-, v_c^+) \in \mathcal{P} \times \mathbb{R} \times \mathbb{R}$, and $\mathcal{P}$ is the set of all environmental processes
- $\mathsf{effect} \in \mathcal{M}$, $\mathcal{M}$ is the set of effects (movement events).
- Delay is a *delay interval* $[d^-, d^+]$, where d^-, d^+ are integers such that $0 \leq d^- \leq d^+$.

In a condition, v_c^- and v_c^+ are the limits of a range within which the value of p_c must fall in order for the condition to hold. The environmental process p_c is drawn from the set of environmental processes $\mathcal{P}$ (which might include processes such as temperature, turbidity, water level, and so forth in the case of fish movements). Note that following Galton (2012), only events can strictly *cause* other events. When one ongoing process is responsible for the continuing operation of another, we refer to this as *perpetuation* rather than cause. Within our general scheme, perpetuation effects can be simulated by dropping the Causes term and using the $\mathsf{Conditions}$ to encode the perpetuating processes. The effect is then an event acting as a proxy for the perpetuated process (as might arise, for example, from discrete observations of what is in reality a continuous process). This results in rules of the form:

$$\mathsf{Conditions} \Rightarrow \mathsf{effect\ after\ Delay}$$

which are the target of the investigations reported here.

2.1 Problem statement

Now assume a data set recording the occurrences of movement events $\mathcal{M}$ at every timestep over some time period $T = [0, n]$ along with the continuously varying values of relevant environmental processes $\mathcal{P}$ across the geographic space in which movement takes place. Our problem is to identify a compact set of rules of the form $\mathsf{Conditions} \Rightarrow \mathsf{effect\ after\ Delay}$ that accurately describe this data.

2.2 Algorithm

Although our full algorithm for dealing with causing events and their interactions with perpetuating processes is rather complex, the component for dealing with perpetuation alone can be relatively simply sketched. Informally, we can:

1. label each time step $t \in T$ as "good" (if the movement effect occurs) or "bad" (if the movement effect does not occur);
2. sort the labeled time steps according to the values of a chosen environmental process p_c;
3. identify consecutive runs of time steps within the sorted sequence; and
4. take the union of the sets of intervals delimiting the start and end points of this each;

In cases where there is only one rule driven by only one perpetuating environmental process, this procedure performs well. Extending the approach to multiple rules (explaining different movement effects), and multiple environmental process (with different conditions), and time delays (resulting from inevitable temporal and spatial granularity effects upon the detection of movement and monitoring environmental processes), makes our full algorithm somewhat more complex. Further, although this algorithm is explicitly temporal, it is not explicitly spatial (as indeed might be expected for a general treatment of causality). Our results also consider the spatial coincidence of perpetuating processes and effects, on the assumption that causal relationships operate in only over immediate temporal and spatial proximity. However, a fuller discussion of the algorithm and these issues is beyond the scope of this extended abstract.

3 Preliminary results and outlook

Our data set involved the set of movement events of more than 1000 tagged fish in the Murray River, south eastern Australia, monitored over five years using a network of 18 logging towers that partition the river system into 24 zones (Koehn et al., 2008). Figure 1 depicts the 24 zones and their downstream adjacencies.

The effects were taken to be the set of fish movements. In our analysis we distinguished upstream and downstream movements, as well as movements between different zones. For example, Figure 2 summarizes the number of downstream movement events between adjacent zones over the five year time period. The Figure shows significantly more movement in later years (most likely a larger-scale effect itself of the decade-long drought in southern Australia which ended in 2010).

Data about water temperature, water level, and salinity from monitoring stations along the river was used as the environmental processes. Space in this extended abstract does not permit an exploration of the full set of results. However, Table 1 gives examples of the best rules found for water level in selected zones, but typical of other zones and processes. The final column in Table 1 shows the F_1 score for these rules. The F_1 score is calculated as the harmonic mean of the precision (positive predictive value) and sensitivity (true positive rate) and provides a useful measure of rule

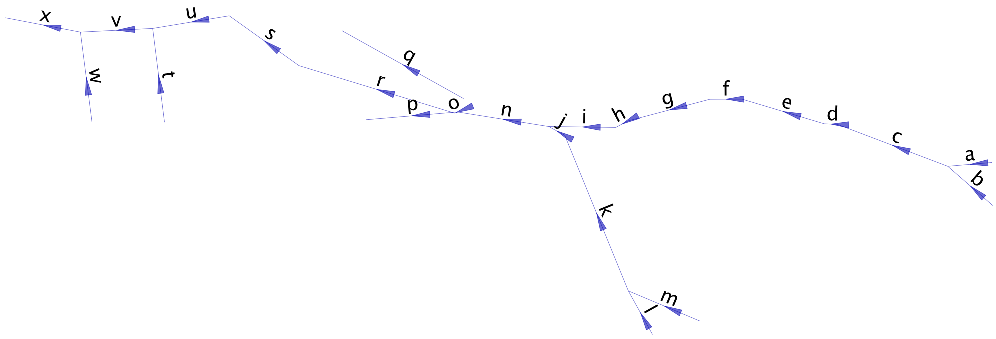

Figure 1: Schematic of zones of monitored river system in the Murray River, Australia, highlighting downstream adjacencies (Koehn et al., 2008). The total length of monitored river is approximately 200km.

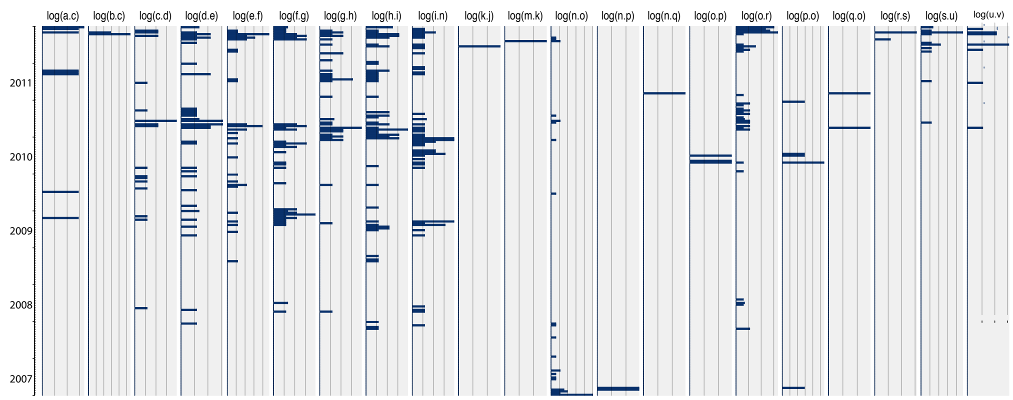

Figure 2: Log of number of fish downstream movements between zones (identified along the columns) over elapsed time.

accuracy (F_1 score of 1.0 indicates no false positives or false negatives, a score of 0.0 indicates only false positives and false negatives). These rules vary between moderately poor performance (e.g., the best rule for zone n, F_1=0.26) up to remarkably good performance. For example, for zone r the rule $1.60 \leq \mathsf{wl} \leq 6.75 \Rightarrow \mathsf{after}\ a$lone can be used to account relatively reliably for movements from r to o within 0 to 5 time steps (F_1=0.62).

Broadly, these results are encouraging in the context of the limitations that 1. these rules only concern the single most effective rule in a zone (additional rules found might further increase the accuracy); 2. the rules only concern water level (other environmental variables might be necessary to explain many fish movements); 3. the rules do not yet account for uncertainty in the data or causal rules; and 4. the rules only concern conditions of perpetuating processes, and do not yet incorporate events (such as moon phase, flood events, or the end of droughts). Current work is investigating each of these limitations.

Zone	Best rule found	F_1 score
i	$126.41 \leq \mathsf{wl} \leq 131.53 \Rightarrow$ moving from i downstream to h after $[0,5]$	0.45
j	$126.89 \leq \mathsf{wl} \leq 126.92 \Rightarrow$ moving from j downstream to k after $[4,5]$	0.26
n	$124.67 \leq \mathsf{wl} \leq 124.75 \Rightarrow$ moving from n downstream to i after $[0,5]$	0.26
o	$1.60 \leq \mathsf{wl} \leq 6.75 \Rightarrow$ moving from o downstream to n after $[0,5]$	0.46
r	$3.02 \leq \mathsf{wl} \leq 6.75 \Rightarrow$ moving from r downstream to o after $[0,5]$	0.62
s	$2.24 \leq \mathsf{wl} \leq 6.40 \Rightarrow$ moving from s downstream to r after $[0,5]$	0.42
u	$2.33 \leq \mathsf{wl} \leq 6.40 \Rightarrow$ moving from u downstream to s after $[0,5]$	0.58
v	$2.78 \leq \mathsf{wl} \leq 6.40 \Rightarrow$ moving from v downstream to u after $[0,5]$	0.48

Table 1: The best rules discovered for selected upstream movement effects, typical of the wider results.

Acknowledgments

AG's research is supported by the UK EPSRC grant EP/M012921/1 "Collectives and Causality". MD's 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." AB's research is supported by funding from the ARC under the Discovery Projects Scheme, project DP120103758 "Artificial intelligence meets wireless sensor networks: Filling the gaps between sensors using spatial reasoning." The authors are also grateful for helpful support and input from the Jarod Lyon and Adrian Kitchingman at the Arthur Rylah Institute (ARI), Melbourne, Australia.

Biography

Antony Galton is Reader in Computer Science at the University of Exeter. His research focuses on spatial and temporal knowledge representation, with applications to artificial intelligence and GI science, including areas such as collective phenomena, and processes and causation in general.

Alan Both is a PhD student at the Department of Infrastructure Engineering, University of Melbourne, Australia. His PhD topic is "Decentralized computation of qualitative spatial relationships in mobile geosensor networks."

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

References

Bleisch, S., Duckham, M., Galton, A., Laube, P., and Lyon, J. (2014). Mining candidate causal relationships in movement patterns. *International Journal of Geographical Information Science*, 28(2):363–382.

Bleisch, S., Duckham, M., Lyon, J., and Laube, P. (2013). Identifying candidate causal relationships in fish movement. In *Proc. GISRUK 2013*, Liverpool, UK.

Buchin, M., Driemel, A., van Kreveld, M., and Sacristán, V. (2011). Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. *Journal of Spatial Information Science*, (3):33–63.

Galton, A. (2012). Formal ontology in information systems. In Donnelly and Guizzardi, editors, *States, Processes and Events, and the Ontology of Causal Relations*, pages 279–292. IOS Press, Amsterdam.

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.

Laube, P. (2014). *Computational Movement Analysis*. Springer.

Laube, P., Imfeld, S., and Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, 19(6):639–668.