Spatial Collectives and Causality

Antony Galton

Department of Mathematics and Computer Science
University of Exeter, UK

GI Forum
Münster, Germany,
27th May 2014

Contents of talk

1. Classifying Collective Motion Patterns
2. States, Events and Causality
3. Mining Candidate Causal Relationships
4. Collective Motion and Causality

PART 1

Classifying Collective Motion Patterns

References

- Zena Wood and Antony Galton, ‘Classifying Collective Motion’ (in Gottfried & Aghajan, eds, *Behaviour Monitoring and Interpretation*, 2009).
- Zena Wood and Antony Galton, ‘Zooming in on Collective Motion’ (in Bhatt et al, eds, Proc. STeDy 2010)

The Three-Level Analysis (TLA)

A full account of the motion of a collective should include components at three levels of **spatial granularity**:

1. **Coarse level:** The motion of the collective as a single entity, as given by the motion of a representative point such as its geometric centroid.
2. **Intermediate level:** The changes to the footprint (as e.g., in Max Dupenois' work)
3. **Fine level:** The motions of the individual members, considered as points.

Temporal Granularity

Fundamental notion is a refinement of the notion of “episode” introduced in the COSIT 2005 paper:

An **episode** (in the refined sense) is a maximal “chunk” of process that looks homogeneous when viewed at a certain granularity.

Here homogeneity is assessed with respect to some set of **qualitative motion descriptors**.

The motion of an individual or collective over an extended period may be regarded as the concatenation of a sequence of episodes, punctuated by **transitions** at which one episode gives way to the next.

A set of qualitative motion descriptors for Level 1

SPEED:

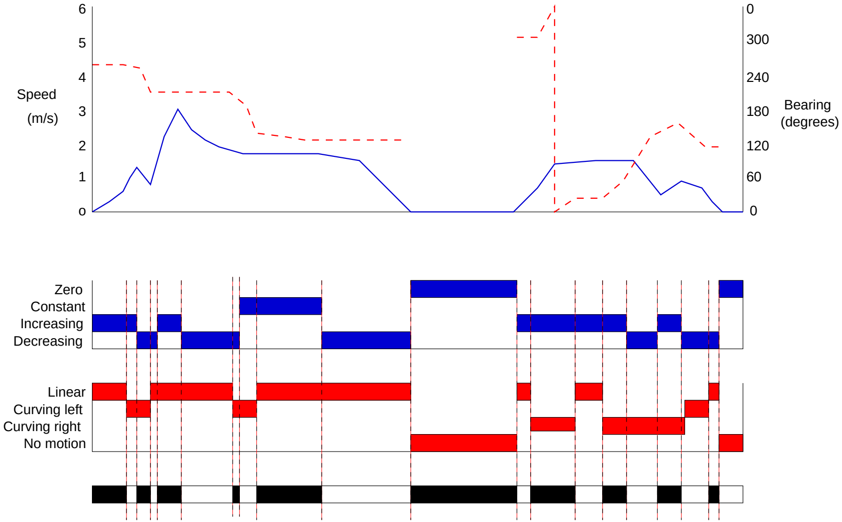
- Zero
- Constant non-zero
- Increasing
- Decreasing

DIRECTION:

- Linear
- Curving left
- Curving right

A more refined set of descriptors might include, for speed, constant, increasing or decreasing acceleration; and for direction, circular, spiralling in, and spiralling out.

Decomposition of motion into qualitative episodes



Qualitative descriptors for Level 2

The chief qualitative characters of a footprint are **size**, **shape**, and **orientation**.

SIZE:

- Constant size
- Expansion
- Contraction

ORIENTATION:

- Constant orientation
- Clockwise rotation
- Anticlockwise rotation

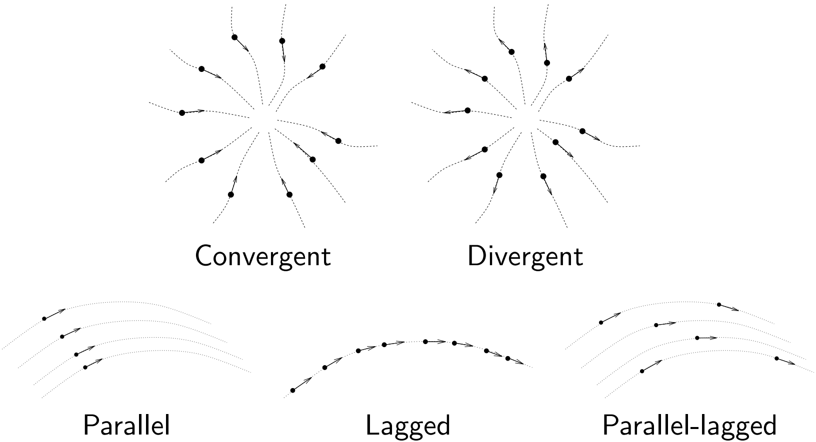
SHAPE — a minefield! There are innumerable dimensions of possible variation, but there has been a lot of work on readily computable and usefully discriminatory shape descriptors.

Qualitative descriptors for Level 3

Here the collective is considered at the granularity level at which the motions of the individual members is apparent. Qualitative descriptors include:

- Uncoordinated
- Convergent
- Divergent
- Parallel
- Lagged
- Parallel-lagged

Five types of coordinated collective motion



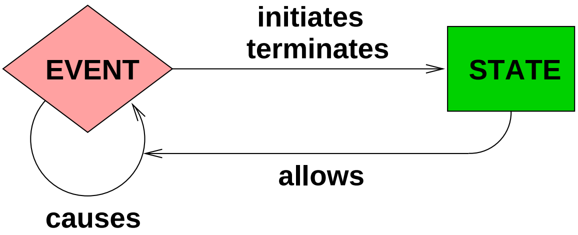
PART 2

States, Events and Causality

References

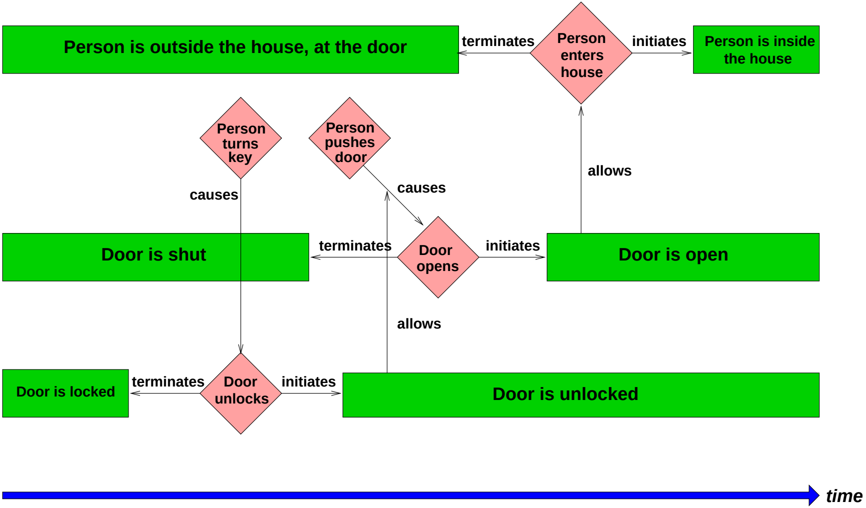
- Antony Galton, 'States, Processes and Events, and the Ontology of Causal Relations', FOIS 2012
- Antony Galton and Mike Worboys, 'Processes and Events in Dynamic Geo-Networks', GeoS 2005

A freezing event INITIATES an iciness state which ALLOWS a braking event to CAUSE an accident. Later, a thawing event TERMINATES the iciness state.

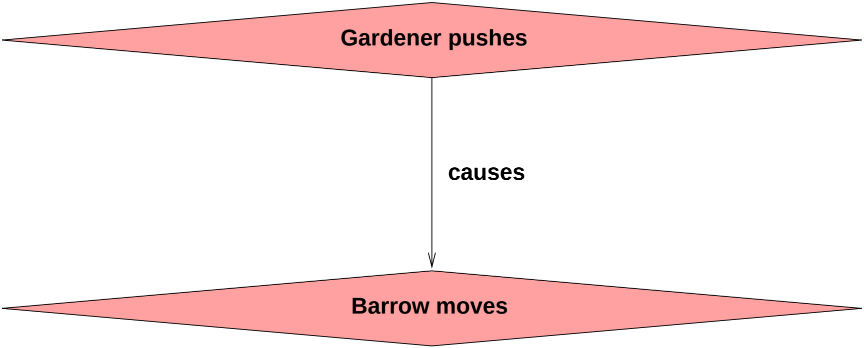


EXAMPLE 1: A person enters a house

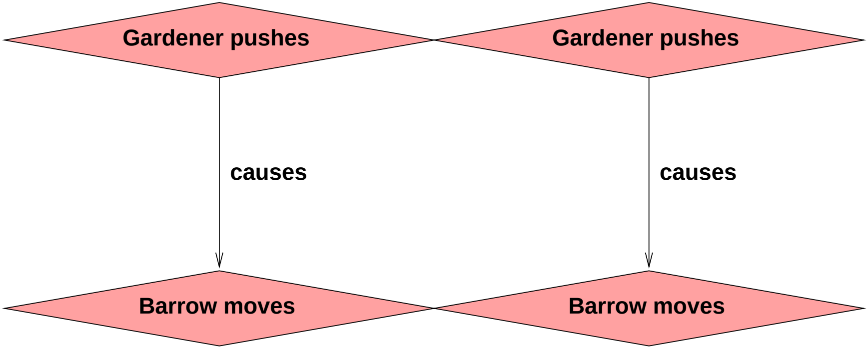
A person is outside a house, at the front door. The door is shut, and locked. The person turns the key, thereby unlocking the door; this allows her to open the door by pushing on it. The result is that the door is then open, which allows her to enter the house by walking forward through the doorway.


Person is outside the house, at the door
terminates
Person enters house
initiates
Person is inside the house
Person turns key
Person pushes door
causes
causes
allows
Door is shut
terminates
Door opens
initiates
Door is open
allows
Door is locked
terminates
Door unlocks
initiates
Door is unlocked
time

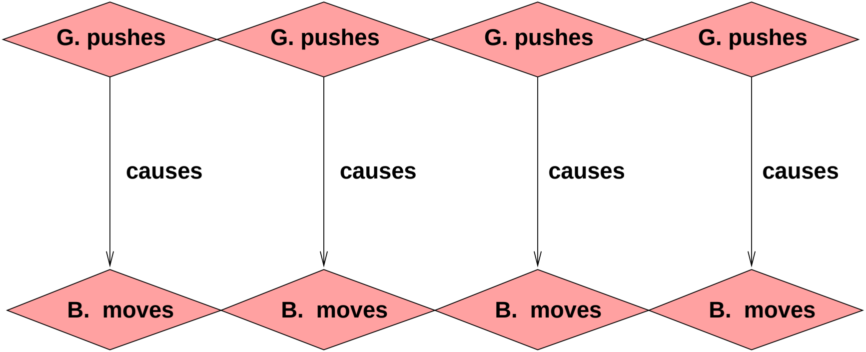
EXAMPLE 2: A gardener pushes a barrow from A to B



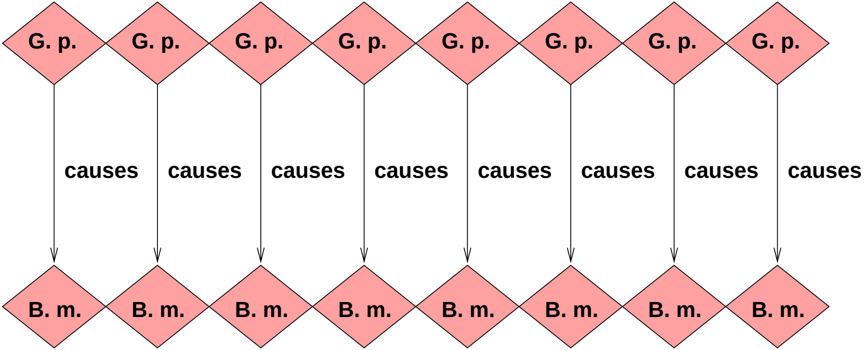
EXAMPLE 2: A gardener pushes a barrow from A to B



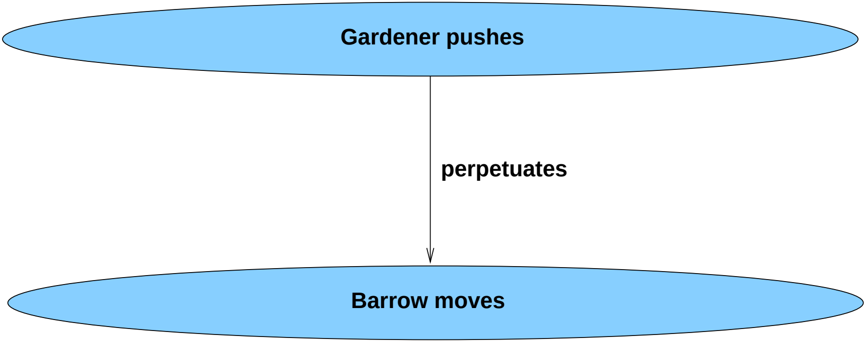
EXAMPLE 2: A gardener pushes a barrow from A to B



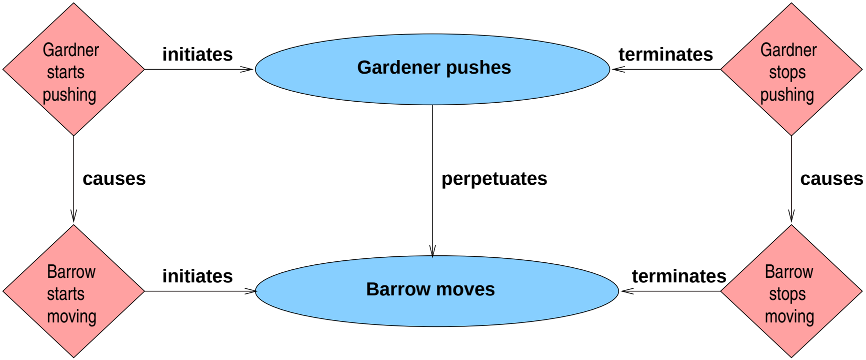
EXAMPLE 2: A gardener pushes a barrow from A to B



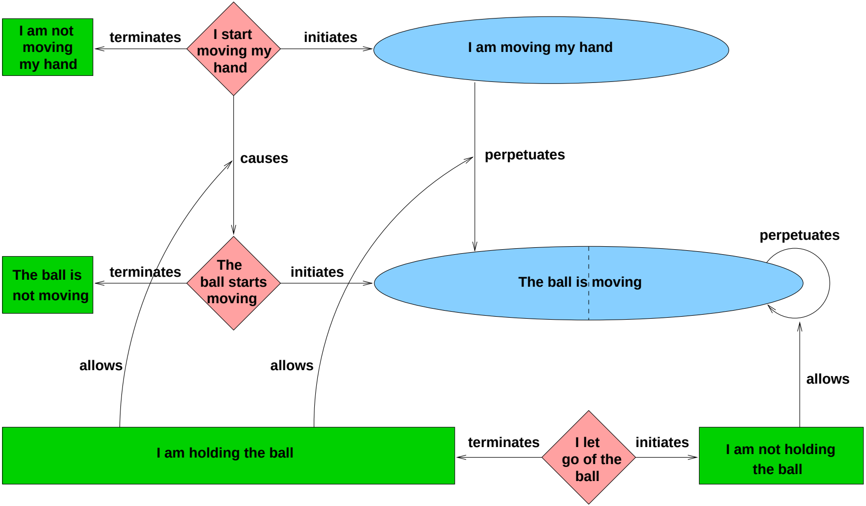
EXAMPLE 2: A gardener pushes a barrow from A to B



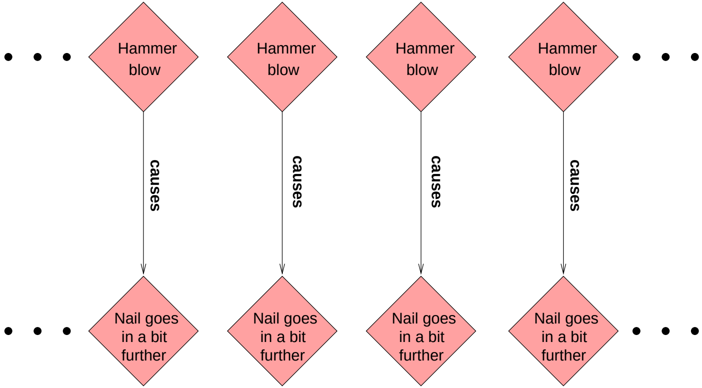
EXAMPLE 2: A gardener pushes a barrow from A to B



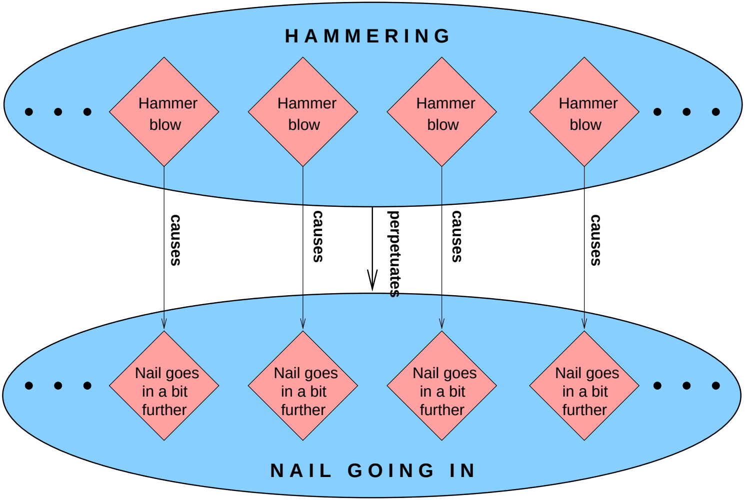
EXAMPLE 3: I throw a ball



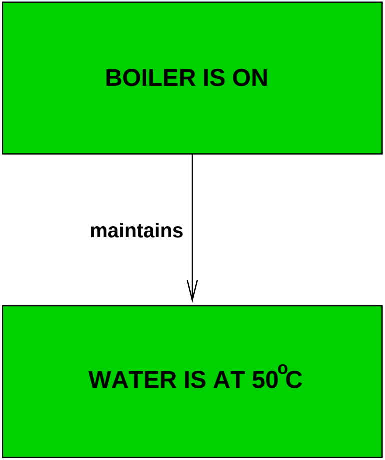
EXAMPLE 4 (Granularity): Hammering in a nail



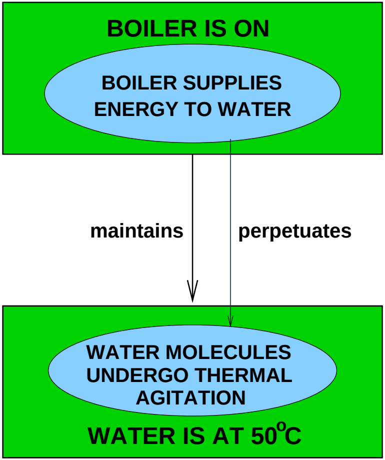
EXAMPLE 4 (Granularity): Hammering in a nail



EXAMPLE 5 (Granularity): Operation of a boiler



EXAMPLE 5 (Granularity): Operation of a boiler



EXAMPLE 5 (Granularity): Operation of a boiler

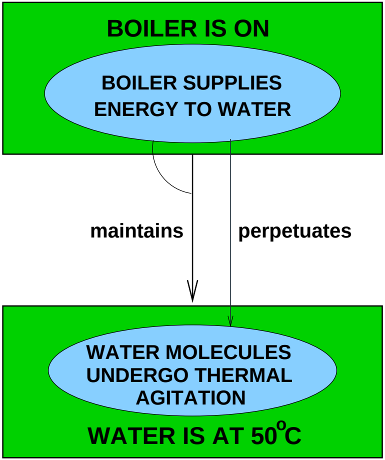
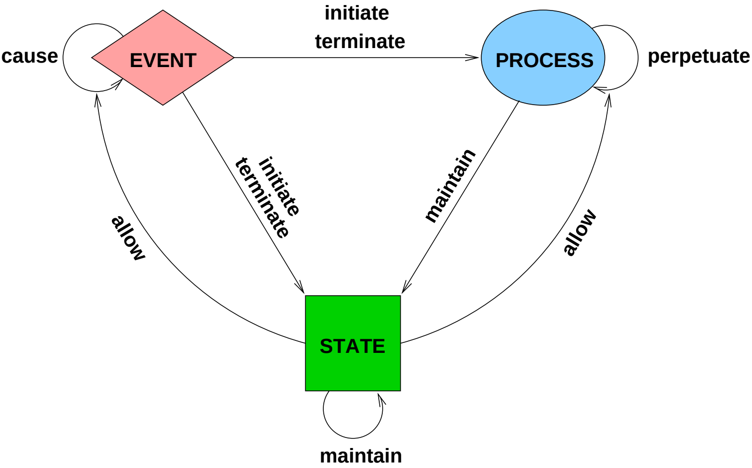


Diagram of Causal and Causal-like Relations



PART 3

Mining Candidate Causal Relationships

Reference

S. Bleisch, M. Duckham, A. Galton, P. Laube, and J. Lyon
'Mining candidate causal relationships in movement patterns'
IJGIS, Volume 28, Number 2, 2014, pp. 363–382.

- Uses *Association Rule Mining* (Agrawal *et al.*, 1993) to look for candidate relationships of the form "*state* allows *event*".
- Uses *Sequence Mining* (Zaki, 2001) to look for candidate relationships of the form "*event* causes *event*".
- Does not handle processes.

The Setting

Lyon (2013) gathered data on fish movement in the Murray River, south-eastern Australia.

- $>$ 1000 fish individuals tagged with radio transmitters.
- 18 river-side radio receivers at strategic locations along river.
- River and its tributaries thereby divided into 24 *zones*.
- Movement of tagged fish between zones tracked over 6 years.
- Environmental states (e.g. water temperature, river flow) and events (e.g., full moon, start of high river flow) also monitored.

The Data: Entities

The data relates to the following sets of entities:

- I, a set of moving-object identifiers
  - tagged fish.
- T, a set of timestamps forming a discrete ordering.
  - days.
- L, a set of locations
  - river zones.
- S, a set of environmental state-types
  - water temperature (five bands), river flow (quartiles)
- E, a set of environmental event-types
  - inception of states, moon phases (quarters).
- M, a set of movement event-types
  - fish movement upstream, downstream, either.

The Data: Relations

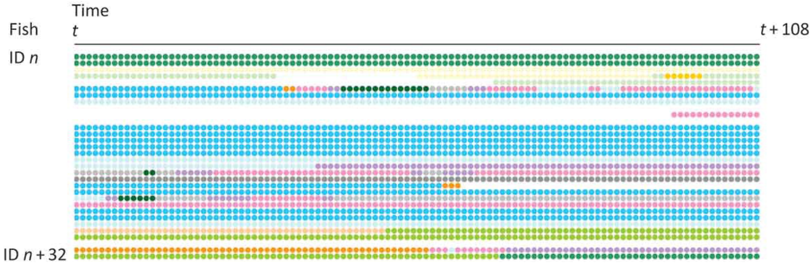
The raw data consist of three sets of triples, as follows:

- $\mathcal{A} \subseteq I \times L \times T$, where $(i, l, t) \in \mathcal{A}$ means individual i is in location l at time t
  - written $At(i, l, t)$
- $\mathcal{H} \subseteq S \times L \times T$, where $(s, l, t) \in \mathcal{H}$ means state s holds in location l at time t
  - written $Holds(s, l, t)$
- $\mathcal{O} \subseteq E \times L \times T$, where $(e, l, t) \in \mathcal{O}$, means that event-type e occurs in location l at time t.
  - written $Occurs(e, l, t)$

In addition the following set of triples is derived from the raw data:

- $\mathcal{P} \subseteq I \times M \times T$, where $(i, m, t) \in \mathcal{P}$ means individual i participates in movement event m at time t.
  - written $Ptp(i, m, t)$

A subset of the raw fish data illustrated



Each horizontal line represents one fish
Each vertical section represents one day
Dot colour indicates river zone in which fish is located on that day

Association Rule Mining: Support and Confidence

For an association rule $m \Rightarrow s$ (where $m \in M, s \in S$), we define

Support is the fraction of (fish,day) pairs for which both m and s are evidenced:

$$\frac{|\{(i,t) \ : \ Ptp(i,m,t) \wedge At(i,l,t) \wedge Holds(s,l,t)\}|}{|I \times T|}$$

Confidence is the fraction of those (fish,day) pairs evidencing m for which s is also evidenced:

$$\frac{|\{(i,t) \ : \ Ptp(i,m,t) \wedge At(i,l,t) \wedge Holds(s,l,t)\}|}{|\{(i,t) \ : \ Ptp(i,m,t)|}$$

Support and confidence are interpreted in the paper as "measures of the strength of evidence that state s 'allows' movement event m to occur".

Sequence Mining

For a sequence $v \rightarrow m$ (where $v \in E, m \in M$), we can define the confidence as the fraction of those occasions on which v occurred that m occurred after a lag of time δt:

$$\frac{|\{(i,t)\ :\ Ptp(i,m,t) \wedge At(i,l,t-\delta t) \wedge Occurs(v,l,t-\delta t)\}|}{|\{(i,t)\ :\ At(i,l,t-\delta t) \wedge Occurs(v,l,t-\delta t)\}|}$$

This is interpreted in the paper as "an indication of the strength of evidence that environmental event v 'caused' movement event m".

Analysis

Three analyses of the data were undertaken:

1. Mining candidate “causes” relationships between environmental and movement events (using $\delta t = 2$ days),
2. Mining candidate “allows” relationships between environmental states and movement events.
3. Mining candidate causal relationships for aggregate movement events (e.g., sequences).

For 1 and 2, the results were validated against the results from the same analyses applied to a simulated data set of randomised fish movement events, using the χ^2 significance test.

Results

For environmental and fish movement events, sequence mining revealed:

- Significant (99% confidence) correlation between initiation of water temperature in the band 20–25°C) and fish movement.
- Significant (95% confidence) correlation between initiation of high river flow and downstream fish movement.
- No significant correlation between moon phase and fish movement.

For environmental states (persisting more than 2 days) and fish movement events, association rule mining revealed:

- Significant (99% confidence) correlation both between water temperature and fish movement and between river flow and fish movement.

These analyses reveal *correlations* in the data; but

Correlation is not Causation!

Inference from correlation to causation depends for its plausibility on an underlying theoretical model.

Can these inferences be made more plausible through a more sophisticated analysis of the kinds of correlation that can result from different kinds of causal relationships?

PART 4

Collective Motion and Causality

Dependencies in collectives

Recall the three levels of collective movement:

1. Movement of the collective as a whole
2. Change in the configuration of the collective
3. Movements of the members of the collective

There are *dependencies* between these levels, some, but not all, of which can be described as causal. Non-causal dependencies include

- *The movement of the collective as a whole is the vector sum of the movements of the individuals*
- *The changes in configuration of the collective are a necessary consequence of all the relative movements of the individuals*

These are mathematical dependencies rather than causal ones.

Causal relations

The causal influences on the movement of each individual in a collective may arise from three different sources:

- From within the individual itself (e.g., an intention to preserve group coherence)
- From amongst other individuals in the collective (e.g., mutual attraction or repulsion)
- From outside the collective (e.g., gravity or other potential gradients, coercion from external agents)

Some causes emanating from outside the collective might act equally on all members (e.g., gravity) and can therefore also be described as acting on the collective as a whole (Level 1 effect).

Others might act selectively, affecting individuals differently from others (e.g. a sieving or filtering process) — these have to be handled at level 3.

How far can the existence of these various kinds of cause be revealed through the Three-Level Analysis of movement patterns?

What I plan to do

In the immediate future:

1. Amass a suitable collection of data-sets within which the different forms of collective causality can be identified.
2. Extend the methods used in Bleisch *et al.* to encompass the full range of causal relations defined in Galton, 2012, taking into account the three-level analysis of movement patterns.
3. Generate and evaluate results of applying the extended methods to the collected data-sets.

In the longer term:

4. Develop an empirically grounded theory of collective causality that can be applied to a wide range of real-life situations.

Thank You for Listening!

ANY QUESTIONS?