Mining Periodic Behaviors for Moving Objects

+ Zhenhui Jessie Li + Bolin Ding + Jiawei Han × Roland Kays ^ Peter Nye

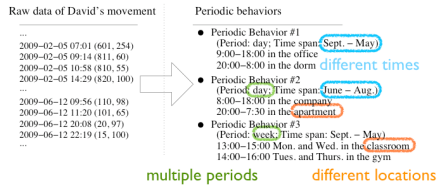
+ University of Illinois at Urbana-Champaign

× New York State Museum

^ New York State Department of Environmental Conservation

Mining periodicity is an important but challenging problem.

Periodicity is a very common phenomenon of moving objects.



Problem: Given a sequence of movement with constant time gap, detect periods and corresponding behaviors.

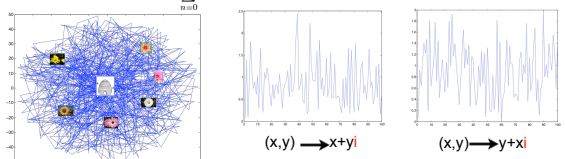
	8:00	9:00	10:00	···	17:00	18:00	19:00
dorm	0.9	0.2	0.1	···	0.2	0.7	0.8
office	0.05	0.7	0.95	···	0.75	0.2	0.1
unknown	0.05	0.1	0.05	···	0.05	0.1	0.1

Table 1: A daily periodic behavior of David.

Directly applying FFT on movement is sensitive to noise.

- Each location (x,y) turns into complex number: x+yi or y+xi
- Apply Discrete Fourier Transform on the sequence

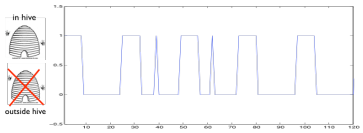
$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn} \qquad k = 0, \ldots, N-1$$



8 hours in hive
6 hours fly nearby

No strong signal indicating the periods. The method is sensitive to the random movement when the bee is outside of the hive.

Finding reference spot(s) to observe the movement is essential to detect the periods.



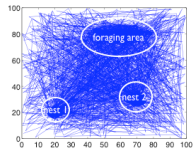
Reference spots: frequently visited locations; should have high density.

The movement is transformed into a binary sequence (in hive or outside hive).

The period in the binary sequence is easier to be detected.

Algorithm Periodica

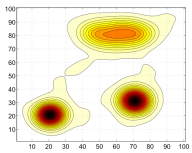
Step 1: Detect periods



Example:

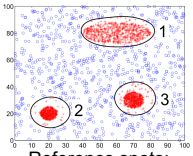
daily behavior 1: nest1&food

daily behavior 2: nest2&food

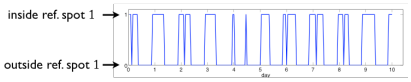


Use kernel-based method to calculate the densities

$$f(c) = \frac{1}{n\gamma^2} \sum_{i=1}^{n} \frac{1}{2\pi} \exp\left(-\frac{|c - loc_i|^2}{2\gamma^2}\right)$$

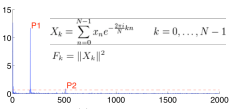


Reference spots: contours of high density places



transform into in-and-out binary sequence

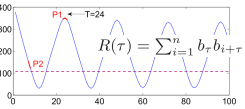
Use Fourier transform & autocorrelation to detect periods



$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn} \qquad k = 0, \ldots, N-1$$

$$F_k = \|X_k\|^2$$

(a) Periodogram



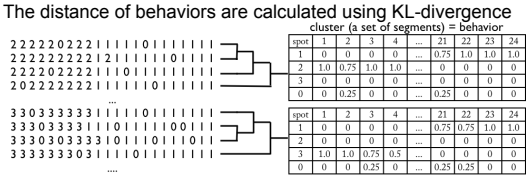
$$R(\tau) = \sum_{i=1}^{n} b_\tau b_{i+\tau}$$

(b) Circular autocorrelation

Step 2: Summarize behaviors

Bottom-up hierarchical clustering

The distance of behaviors are calculated using KL-divergence



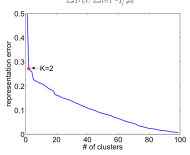
cluster (a set of segments) = behavior

spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	1.0	1.0	1.0
2	1.0	0.75	1.0	1.0	...	0	0	0	0
3	0	0	0	0	...	0	0	0	0
0	0	0.25	0	0	...	0.25	0	0	0

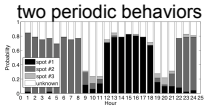
spot	1	2	3	4	...	21	22	23	24
1	0	0	0	0	...	0.75	0.75	1.0	1.0
2	0	0	0	0	...	0	0	0	0
3	1.0	1.0	0.75	0.5	...	0	0	0	0
0	0	0	0.25	0	...	0.25	0.25	0	0

representation error to determine # of behaviors

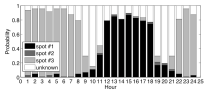
$$E(C) = \frac{\sum_{l \in C} \sum_{i=1}^{\tau} \mathbf{1}_{l'_i \neq 0} \cdot (1 - p(x_i = l'_i))}{\sum_{l \in C} \sum_{i=1}^{\tau} \mathbf{1}_{l'_i \neq 0}}$$



two periodic behaviors



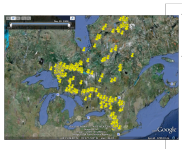
(a) $\mathbf{P}$ of periodic behavior #1



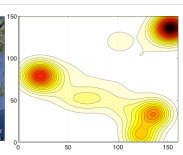
(b) $\mathbf{P}$ of periodic behavior #2

Experiments

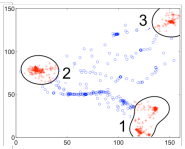
A real bald eagle movement case



Original data (3 years)



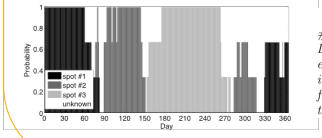
density map



reference spots

spot	T
1	363
2	363
3	364

periods detected



"This bald eagle stays in New York area (i.e., reference spot # 1) from December to March. In March, it flies to Great Lakes area (i.e., reference spot #2) and stays there until the end of May. It flies to Quebec area (i.e., reference spot #3) in the summer and stays there until late September. Then it flies back to Great Lake again staying there from mid October to mid November and goes back to New York in December."

A synthetic human movement case

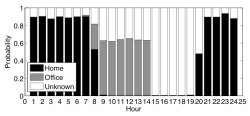
data

20:00~8:00 at home

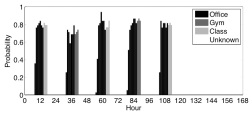
9:00~14:00 at office on weekdays

15:00~17:00 at gym on Tues. &Thurs.

15:00~17:00 at class on Mon., Wed., & Fri.

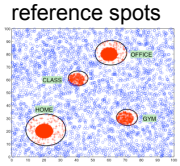


(a) Periodic behavior for $T = 24$.



(b) Periodic behavior for $T = 168$.

reference spots



periods detected

spot	period
Home	24
Office	24, 168
Gym	168
Class	168