

# Preliminary investigation of co-occurrence rule extraction from multi-modal animal locomotion data

田一鳴 1) 前川卓也 1) 天方大地 1) 原隆浩 1) 松本祥子 2) 依田憲 2)

藤岡慧明 3) 濱井郁弥 4) 福井大 5) 飛龍志津子 6)

1 大阪大学大学院情報科学研究科 2 名古屋大学大学院環境学研究科

3 同志社大学研究開発推進機構 4 同志社大学大学院生命医科学研究科

5 東京大学大学院農学生命科学研究科 6 同志社大学生命医科学部

## 1. はじめに

動物の行動情報をセンサデバイスを用いて記録する手法はバイオロギングと呼ばれ[1], 位置座標(軌跡)のみならず, 環境データや生体データなどの様々な時系列データをセンサにより同時に計測することで, 多くの研究者が動物の移動メカニズムの解明を目指している. すなわち, 動物のデータの移動の解析には, 軌跡情報のみでなく, 同時に計測された時系列センサデータを同時に解析して, 知識抽出を行う必要がある.

一般的に動物の移動は複数のモードから構成される. 例えば, 餌などの探索を局所的に行う「局所探索」や, 餌場間などの移動を行う「長距離移動」などから構成されている. 一方で, 環境センサや生体センサから得られたデータにもモードは存在し, 例えば脳波の時系列データから, 動物の状態を「リラックス」モードと「緊張」モードに分けることもできる. 動物から得られた複数のモーダルな時系列センサデータ(軌跡, 環境データ, 生体データなど)をそれぞれこのようなモードにセグメントし, 異なるモーダルを跨いだ頻出パターンを抽出することで, 生態学者にとって異なるモード間の相互関係を理解しやすい有用な知見を提供できる.

このような, 時系列データ(軌跡データやセンサデータ)からのモードの抽出には, 系列データクラスタリング手法が用いられてきた. しかし, これまでの手法では, クラスタリングに用いる特徴量や, 閾値などのパラメータは, 研究者の先験的な知見から主観的に決められていた.

そこで, 本研究ではそのような特徴量や閾値を自動的に決定し, 頻出ルールの抽出を行う手法を開発する. ここで, 本研究が対象とする動物のデータからの頻出ルール抽出の最終目標は, 有用度の高い頻出ルールを見つけることであるため, 得られる頻出ルールの有用度が最も高くなるようなパラメータを同時に発見する. すなわち,

それぞれの系列データのある特徴量とクラスタリングパラメータを用いてクラスタリングし, それらの結果から頻出ルールの抽出およびその有用度の計算を行う. この手順を, 利用する特徴量とパラメータを変えながら何度も行い, 高い有用度が得られたモード分けを, 尤もらしいモード分けであるとする.

しかし, 軌跡データや時系列データから抽出される特徴量は数多く存在し, クラスタリングパラメータも様々な値を取り得る. そのため, 時系列データのクラスタリングを何度も繰り返し行う必要があり, その計算量は膨大となる. そこで本研究では, パラメータを変更して何度も繰り返すクラスタリング処理を高速化して頻出ルールを導出する手法を提案する.

## 2. 提案手法

### 2.1 提案手法概要

提案手法は, 移動軌跡 $P$ (位置座標の時系列データ)と任意の次元数の時系列データ $S$ (環境センサデータなど)を入力とする. また, これらの時系列データは同時に記録されたものとする. まず, それぞれの時系列データから特徴を抽出して, 複数の1次元の特徴時系列データを計算する.  $P$ からは $P_0, P_1, \dots, P_N$ の特徴時系列データが,  $S$ からは $S_0, S_1, \dots, S_M$ が得られる. そして, それぞれの $P_n$ と $S_m$ のペアについて, ルールとそのスコアを導出する.  $P_n$ と $S_m$ それぞれに対して時系列クラスタリングを行うと, パラメータごとに時系列クラスタリング結果 $P_{n,\theta_0}, P_{n,\theta_1}, \dots, P_{n,\theta_j}, \dots$ と $S_{m,\theta_0}, S_{m,\theta_1}, \dots, S_{m,\theta_j}, \dots$ が得られる.  $P_{n,\theta_i}$ と $S_{m,\theta_j}$ の組み合わせごとに, ルールとその有用度(スコア)のペア群を導出する. 有用度は,  $P_{n,\theta_i}$ と $S_{m,\theta_j}$ のあるクラスタ同士の共起度に基づいて計算される.

### 2.2 時系列クラスタリング

得られた $P_n$ (もしくは $S_m$ )に対して, 与えられたパラメータごとに網羅的に時系列クラスタリングを行う. 時系列クラスタリングは, (1)時系列セグメンテーションと(2)スペクトラルクラスタリングの2つの手順に分けられる.

#### (1) 時系列セグメンテーション

まず,  $P_n$ に含まれる全データポイントの平均と分散を計算したあと, SAX[2]を用いて各データポイントのシンボル化を行う. 次に, (2)スペクトラルクラスタリングが必要となる, 各シンボルの代表値(平均)を求める.

#### (2) スペクトラルクラスタリング

次に, 時系列セグメンテーションの結果のクラスタリ

1 Graduate School of Information Science and Technology Osaka University

2 Graduate School of Environmental Studies Nagoya University

3 Organization for Research Ubuitatues and Development Doshisha University

4 Graduate School of Life and Medical Sciences Doshisha University

5 Graduate School of Agricultural and Life Sciences, The University of Tokyo

6 Doshisha University Faculty of Life and Medical Sciences

ングを、グラフ分割手法の1つであるスペクトルクラスタリング[3]を用いて行う。ただし、長時間の時系列データの場合、ノード数が膨大となるため、グラフのノードの削減を行うことでスペクトルクラスタリングの計算時間を削減する。また、さらに計算時間を削減するため、この手順を以下のように手順2-1と手順2-2に分ける。

手順2-1. まず、時系列セグメンテーションの結果からグラフを作成する。提案手法では、同じシンボルをもつノードを1つに併合することで、ノード数の削減を行う。このノード数を削減したグラフを用いてスペクトルクラスタリングを行う。スペクトルクラスタリングでは、ノード間の類似度を示すラプラシアン行列を計算したあと、その行列の固有値分解を行うことで、固有ベクトルと固有値を求める。

手順2-2. 上記手順で求めた固有ベクトルを基に、k-means++法[4]を用いてノードのクラスタリングを行う。

#### 4. 評価実験

提案手法の評価を、新潟県岩船郡粟島に生息するオオミズナギドリから得られたGPSと水深センサデータ

[Matsumoto et al. 2017]を用いて行う。オオミズナギドリは、飛行中に海中へ飛び込むため、水深センサデータを取得している。これらのデータを提案手法を用いて分析することで、飛び込み行動と移動モードの関係性に関するルールが抽出されることが期待される。

移動軌跡からは、速度[m/s]、加速度とその絶対値[m/s<sup>2</sup>]、角速度[rad/s]、速度の移動平均・移動分散、加速度の移動平均・移動分散、加速度の絶対値の移動平均・移動分散、角速度の移動平均・移動分散を計算した。移動平均・分散の窓幅は10, 20, 30, 40, 50分に設定してそれぞれ計算した。水深センサデータに対して微分を計算したあと、絶対値を計算した。

上記の手法をオオミズナギドリのデータセットに適用した際の計算時間は13.41秒であった。データセットから得られた最も有用性の高かったルールは、「加速度の絶対値の移動分散の時系列」と「深度センサの微分の絶対値の時系列」のモード分けから得られたものであった。

「加速度の絶対値の移動分散の時系列」から得られた、「加速度の絶対値の移動分散が大きいモード(約0.5(m/s<sup>2</sup>)<sup>2</sup>以上)」と「深度センサの微分の絶対値が大きいモード(約1Pa/s以上)」がよく共起していた。加速度の絶対値の移動分散から得られたモードは「速度変化が大きい」、深度センサの微分の絶対値から得られたモードは「海面に飛び込むモード」と言うことができる。図1に、それぞれのモードに対応するセグメントを赤色でハイライトした、ある個体から収集された軌跡データを示す。また、その際のそれぞれの時系列データも示す。赤色でハイライトされたデータ点が、上記のモードに対応するデータ点である。

#### 5. おわりに

本稿では、移動体から得られたマルチモーダル時系列データから高速に頻出ルールを抽出する手法の提案を行った。これまでの時系列データからのクラスタリングでは、主観的な特徴およびパラメータ選択が行われていたが、提案手法ではルール抽出と 동시에特徴およびパラメータ選択も同時に行う。このとき、適切な特徴およびパラメータ選択を発見するため、高速にクラスタリングを繰り返しつつルール抽出を行う。今後は、それぞれの動物から得られた頻出ルールの生物学的意味の詳細な検証を行う予定である。

謝辞：本研究は、JSPS 科研費 JP16H06539, JP16H06542, JP16H06541 の助成を受けて行われたものである。

#### 参考文献

- [1] Rutz, C. and Hays, G. C.: New frontiers in biologging science (2009).
- [2] Lin, J., Keogh, E., Wei, L. and Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series, *Data Mining and knowledge discovery*, Vol. 15, No. 2, pp. 107- 144 (2007).
- [3] Ng, A. Y., Jordan, M. I. and Weiss, Y.: On spectral clustering: Analysis and an algorithm, *Advances in neural information processing systems*, pp. 849-856 (2002).
- [4] Arthur, D. and Vassilvitskii, S.: k-means++: The advantages of careful seeding, *ACM-SIAM symposium on Discrete algorithms (SODA)*, pp. 1027-1035 (2007).

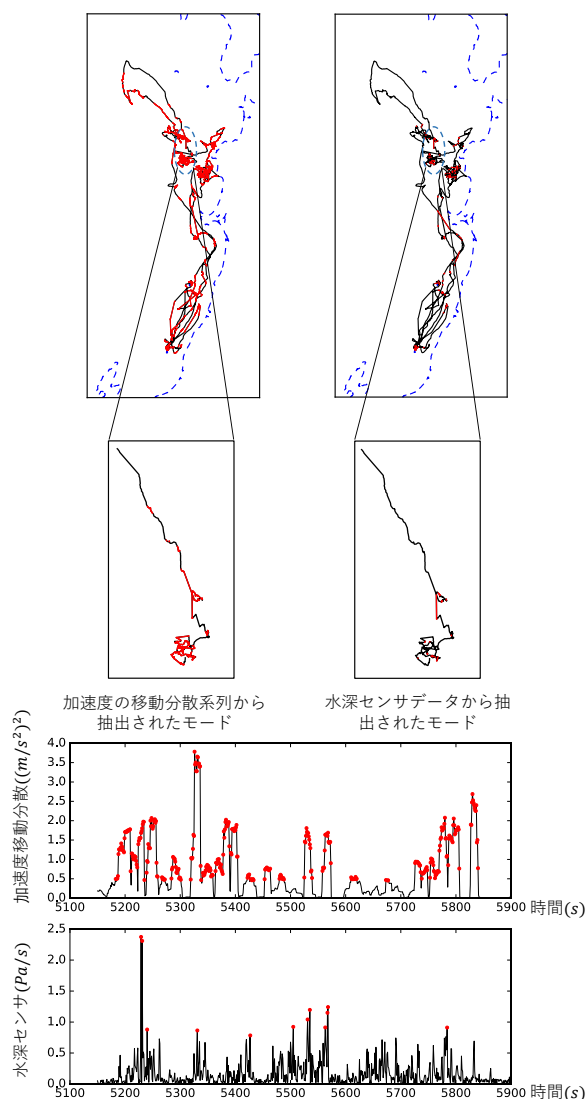


図1 オオミズナギドリのデータから得られたルールの例