Animal Biotelemetry

METHODOLOGY

Open Access

Automatic identification of differences in behavioral co-occurrence between groups

Yiming Tian[1], Takuya Maekawa[1*], Joseph Korpela[1], Daichi Amagata[1], Takahiro Hara[1], Sakiko Matsumoto[2] and Ken Yoda[2]

Abstract

Background: Recent advances in sensing technologies have enabled us to attach small loggers to animals in their natural habitat. It allows measurement of the animals' behavior, along with associated environmental and physiological data and to unravel the adaptive significance of the behavior. However, because animal-borne loggers can now record multi-dimensional (here defined as multimodal) time series information from a variety of sensors, it is becoming increasingly difficult to identify biologically important patterns hidden in the high-dimensional long-term data. In particular, it is important to identify co-occurrences of several behavioral modes recorded by different sensors in order to understand an internal hidden state of an animal because the observed behavioral modes are reflected by the hidden state. This study proposed a method for automatically detecting co-occurrence of behavioral modes that differs between two groups (e.g., males vs. females) from multimodal time-series sensor data. The proposed method first extracted behavioral modes from time-series data (e.g., resting and cruising modes in GPS trajectories or relaxed and stressed modes in heart rates) and then identified two different behavioral modes that were frequently co-occur (e.g., co-occurrence of the cruising mode and relaxed mode). Finally, behavioral modes that differ between the two groups in terms of the frequency of co-occurrence were identified.

Results: We demonstrated the effectiveness of our method using animal-locomotion data collected from male and female Streaked Shearwaters by showing co-occurrences of locomotion modes and diving behavior recorded by GPS and water-depth sensors. For example, we found that the behavioral mode of high-speed locomotion and that of multiple dives into the sea were highly correlated in male seabirds. In addition, compared to the naive method, the proposed method reduced the computation costs by about 99.9%.

Conclusion: Because our method can automatically mine meaningful behavioral modes from multimodal time-series data, it can be potentially applied to analyzing co-occurrences of locomotion modes and behavioral modes from various environmental and physiological data.

Keywords: Animal behavior, Co-occurrence of behavioral modes, Time-series, Trajectory

Background

Animal behavior comprises a continuous stream of events that can be divided into discrete behavioral modes (e.g., resting and foraging), although they have generally been defined in various terms such as behavioral repertories, units, and categories [1, 2]. These behavioral modes are measured for each sensor modality (i.e., different sensors used when observing behaviors and environments such as sounds and locations). Some behavioral modes obtained from different sensor modalities may co-occur (we term "behavioral co-occurrence" in this paper), and the co-occurrence probability of the different behavioral modes depends on different groups or

*Correspondence: maekawa@ist.osaka-u.ac.jp
[1] Graduate School of Information Science and Technology Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
Full list of author information is available at the end of the article

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

environments. For example, chimpanzees call while they are foraging, and the co-occurrence probability of the call and foraging behaviors depends largely on sex and individual situations [3]. Behavioral co-occurrences reflect the decision-making of each group, which provides an understanding of sex-specific foraging strategies, polymorphism with the corresponding strategies, and the life-history strategies of different populations. Therefore, it is important to identify behavioral co-occurrences and their differences between groups from large behavioral data to understand animal behavior and/or its relationship with surrounding environments.

Recent advances in sensing technologies have enabled the attachment of small loggers, allowing remote measurement of animal behavior along with associated environmental and physiological data and elucidation of the adaptive significance of the behavior [4]. Sensor data from each modality (such as GPS trajectories and heart rates) comprises several behavioral modes, including resting and cruising movements and relaxed and stressed physiological states, as shown in the hypothetical example (Fig. 1). As behavioral modes measured by each sensor modality are a reflection of animals' internal hidden states (e.g., motivation), extracting frequent co-occurrences of behavioral modes is crucial to estimating these hidden states and understanding group differences. The example of frequent co-occurrences of behavioral modes in Fig. 1 shows that behavioral mode 1-1 and behavioral mode 2-1 appear concurrently in seabirds. Assuming that behavioral mode 1-1 is identified as the cruising mode and behavioral mode 2-1 is identified as the relaxing mode, this co-occurrence indicates that cruising flights are more likely to contribute to energy saving during movement rather than food searching flights.

In this study, we develop a method that detects co-occurrence of behavioral modes that differ in frequency between two groups (e.g., males and females). We developed an effective algorithm to extract co-occurrences of behavioral modes from time-series data, as shown in Fig. 1. Although Fig. 1 is a simple example easily verified visually or using a simple method, it is often difficult to identify co-occurrences of several behavioral modes from multimodal data in terms of difficulties in (i) behavioral mode detection from multimodal data and (ii) parameter selection. Supervised and unsupervised methods have been used to detect behavioral modes from time-series behavioral data. The supervised methods, such as Random Forest, are trained on labeled data and predict a class label for each data point at a time slice [5, 6]. As it is difficult to collect labeled data from wild animals, unsupervised methods such as clustering methods and hidden Markov models (HMMs) have been studied actively [7, 8]. However, processing multimodal data using unsupervised methods is problematic. In the unsupervised methods, data are grouped based on the distance between the data points, which makes it difficult to properly cluster data points in multimodal sensor data. Assume that a data point is composed of an acceleration value and an ambient temperature value. In this case, even when an animal performs the same activity, which is reflected in the acceleration data, the data points collected in the high-temperature condition can be grouped into the different cluster from those collected in the low-temperature condition; this makes it difficult to interpret such meaningless clusters. Although Adam et al. [9] employed hierarchical hidden Markov models (HHMMs) to estimate the latent states for each sensor modality, the latent states of the sensors are dominated by those of the main sensor, which is specified by a researcher. Therefore, prior studies have only been able to handle multimodal data with high correlations, for example, multivariate data composed of displacement and accelerometer data.

In addition to the aforementioned problem, prior supervised and unsupervised behavioral mode detection methods have relied on manually defined variables of behavior (called a "feature" in machine learning), such as locomotion speed [10], with the predetermined number of behavioral modes (called a clustering "parameter" in the unsupervised methods). Although results of behavioral mode detection are affected greatly by the selection of features and parameters by a researcher, it is impossible to manually select a combination of features and parameters that yields meaningful behavioral modes from high-dimensional time-series data. To address these issues, we developed a computationally efficient method to detect the frequent co-occurrence of behavioral modes by automatically selecting the parameters and features that maximize the usefulness of the obtained co-occurrence of those behavioral modes (computed based on the correlation between the two behavioral modes). In addition, we use deep learning to automatically extract meaningful features from behavioral data.

The contributions of this study are as follows: (1) We propose a method to automatically detect frequent co-occurrence of behavioral modes that differ in frequency between two groups from multimodal animal behavioral data. Our method find behavioral modes from different sensor modalities that frequently co-occur, and then identifies behavioral modes that differ between the two groups in terms of the frequency of co-occurrence. To our knowledge, this is the first study to automatically detect knowledge related to the co-occurrence of behavioral modes from multimodal animal behavioral data. (2) We develop a computationally effective method that finds optimum hyperparameters (combinations of features used and clustering parameters). (3) We validate the

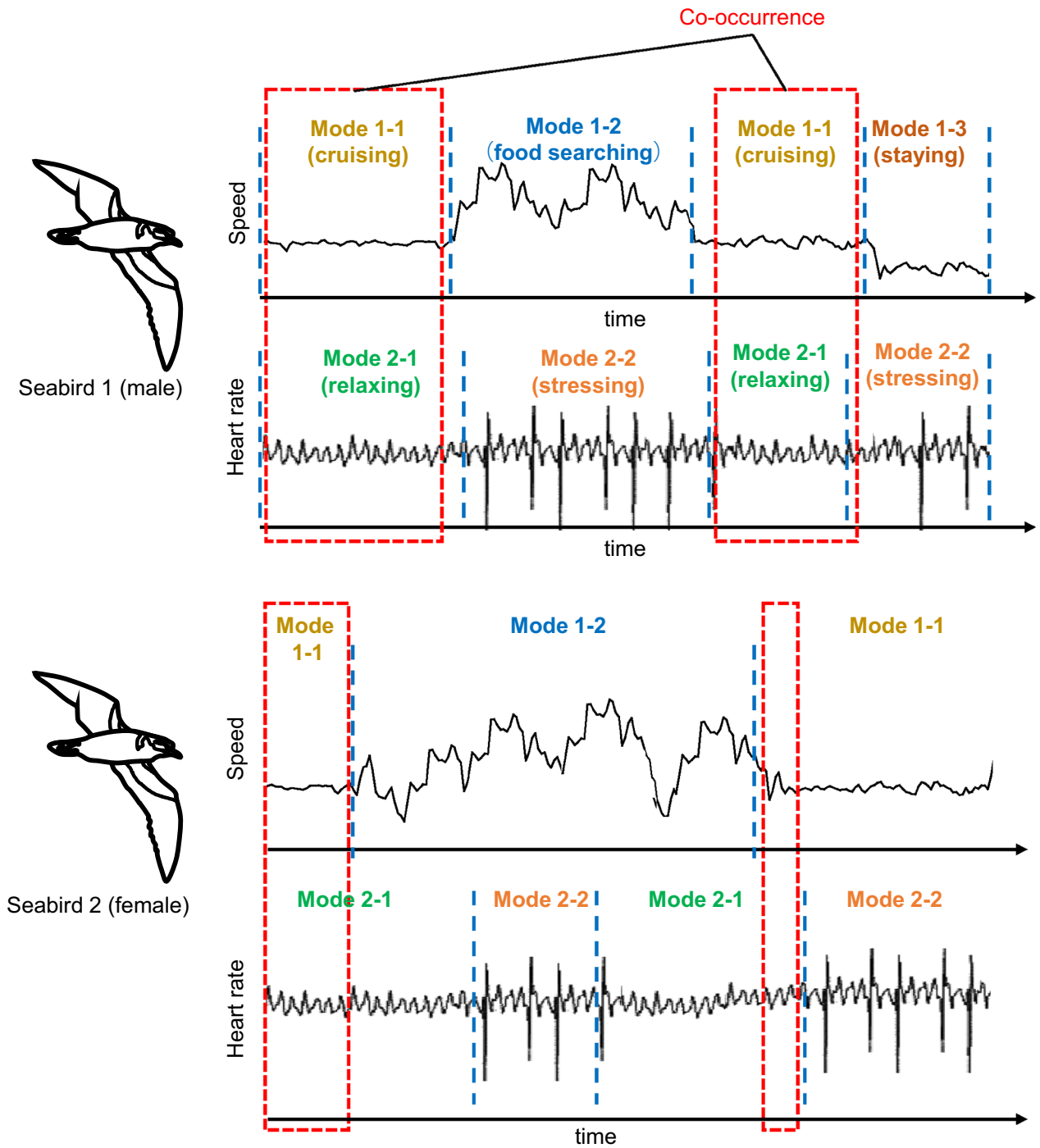

Fig. 1 A hypothetical example of co-occurrence of behavioral modes. Behavioral mode 1-1 extracted from the time-series speed data and behavioral mode 2-1 extracted from the time-series heart rate data co-occur frequently. In this example, the meanings of behavioral modes 1-1 and 2-1 are translated as the cruising mode and relaxing mode based on their respective value ranges. This example also indicates that the frequency of the co-occurrences for the male seabird is higher than that for the female seabird

effectiveness of the method using actual data obtained from seabirds, including sex-specific differences in seabirds related to co-occurrences between horizontal (i.e., GPS positions) and vertical movements (i.e., diving depths).

Materials and methods

Our method comprised of four steps (Fig. 2): (1) feature extraction/learning, (2) behavioral mode extraction using a clustering method, (3) identifying co-occurrence behavioral modes, and (4) identifying behavioral co-occurrence differences between groups. The method started with original sensor data (e.g., GPS coordinates) and calculated or learnt features (e.g., speed) in the first step. Then, the method extracted behavioral modes (e.g., cruising mode and relaxing mode) from the time-series of features in the second step. In the third step, the method identified pairs of behavioral modes that frequently co-occur. In the fourth step, the method output pairs of behavioral modes that differ between given two groups in terms of the frequency of co-occurrence.

Time-series segmentation and clustering methods have been used in previous studies to extract behavioral modes from behavioral data such as trajectories and time-series sensor data [11]. The time series of feature values (e.g., the time series of movement speeds) were extracted from the original time-series data and then segmented and clustered to identify behavioral modes for each feature. In previous studies, the features extracted from the original time-series data and the parameters of the methods (e.g., threshold and expected number of behavioral modes) were manually selected and/or designed based on their experience regarding to the study species [12, 13]. By contrast, the proposed method could identify co-occurrences of behavioral modes by changing parameters and features without a priori assumptions regarding the threshold and number of behavioral modes. However, this required a large number of combinations for testing, resulting in high computational costs. To achieve this in reasonable time, we proposed a method that reduces the total computational cost.

Problem definition

Our objectives were as follows: (1) find highly correlated behavioral modes between data with multiple time series (e.g., GPS data, water pressure data, and ambient temperature data); and (2) identify co-occurrence of behavioral modes that differ between groups.

Let L and E be a location-information dataset and a dataset of other sensors (e.g., acceleration sensor, water pressure sensor, and temperature sensor including environmental sensor and physiological sensor), respectively. Specifically, $L = \{\langle l_1, t_1\rangle, ..., \langle l_{|L|}, t_{|L|}\rangle\}$, where l_i is two-dimensional (2D) (or three-dimensional) location information, and t_i is the time when l_i is observed. Additionally, $E = \{\langle e_1, t_1\rangle, ..., \langle e_{|E|}, t_{|E|}\rangle\}$, where e_i represents d-dimensional sensor data, and d is $\geqslant 1$. For example, when we observe atmospheric pressure and body temperature, $d = 2$. Our inputs were obtained from L and E. Given L, we could extract multiple types of one-dimensional-feature time series, L_1, L_2, ..., L_n, comprising handcrafted features and features learned using feature-learning approaches [14].

For example, L_1 is a time series of speed. Note that, when data from multiple individuals were given, L_i was created by concatenating time-series of a corresponding feature from the individuals. Furthermore, given E, E_1, E_2, ..., and E_m were obtained, where E_1 might be a time series of the derivative of atmospheric pressure. Next, we defined several terms related to clustering.

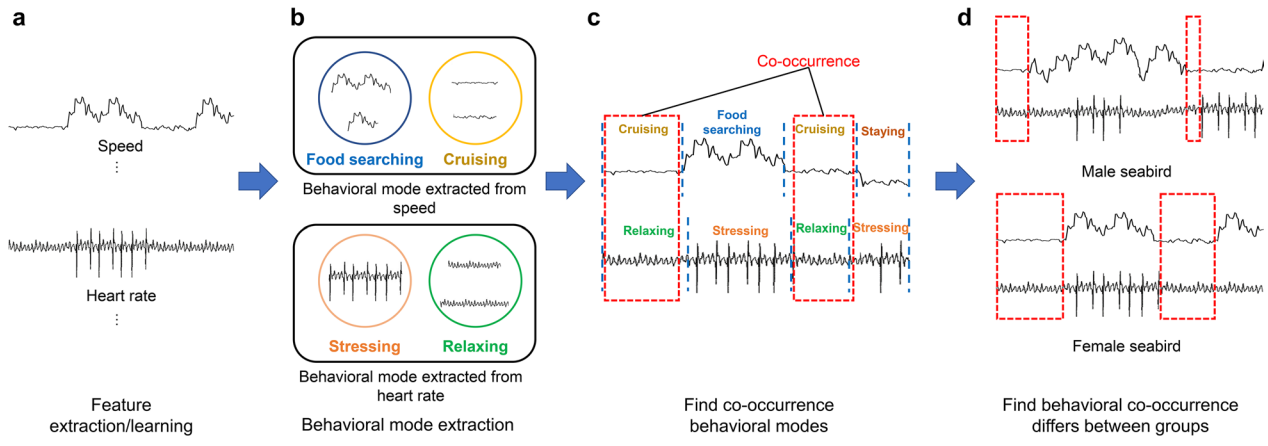

Fig. 2 Overview of the proposed method. **a** Feature extraction/learning. **b** Behavioral mode extraction using a clustering method. **c** Identifying co-occurrence of behavioral modes. **d** Identifying behavioral co-occurrence differences between groups

Definition 1 (Segment). Given a time series X and a parameter α, a segment s that is computed based on α is a subsequence of X. If $s = \{\langle x_a, t_a\rangle, \langle x_{a+1}, t_{a+1}\rangle, ..., \langle x_b, t_b\rangle\}$, the time length of s, denoted by $s.t$, is $[t_a, t_b]$.

Definition 2 (Cluster). Given a time series X and a parameter set θ comprising α and β, a cluster that is computed based on β is a set of segments that follows Definition 1. Each cluster can be referred to a behavioral mode.

Given L_i, E_j, and clustering parameter sets θ_n and θ_m, we have L_{i,θ_n} and E_{j,θ_m}, which are respectively the clustering results of L_i and E_j. The clustering result L_{i,θ_n} contains clusters, and each cluster (i.e., behavioral mode) is composed of segments belonging to the behavioral mode. Let L^p_{i,θ_n} (E^q_{j,θ_m}) be the p-th (q-th) cluster (i.e., the behavioral mode) in L_{i,θ_n} (E_{j,θ_m}). Note that p (q) is larger than or equal to 1 and smaller than or equal to the number of clusters. Here, we defined the score of the pair of L^p_{i,θ_n} and E^q_{j,θ_m}.

Definition 3 (Score($L^p_{i,\theta_n}, E^q_{j,\theta_m}$)). Given $L^p_{i,\theta_n} = \{s^p_1, s^p_2, ...\}$ and $E^q_{j,\theta_m} = \{s^q_1, s^q_2, ...\}$,

$$Score(L^p_{i,\theta_n}, E^q_{j,\theta_m}) = Corr(L^p_{i,\theta_n}, E^q_{j,\theta_m}) - Freq(L^p_{i,\theta_n}, E^q_{j,\theta_m}),$$

where

$$Corr(L^p_{i,\theta_n}, E^q_{j,\theta_m}) = \frac{\sum_{n,m} |s^p_n.t \cap s^q_m.t|}{\sum_m |s^q_m.t|}$$

and

$$Freq(L^p_{i,\theta_n}, E^q_{j,\theta_m}) = C\frac{\max(\sum_n |s^p_n.t|, \sum_m |s^q_m.t|)}{|L_i|}.$$

Note that $Corr(L^p_{i,\theta_n}, E^q_{j,\theta_m})$ measures the time correlation between L^p_{i,θ_n} and E^q_{j,θ_m} (i.e., co-occurrence between a behavioral mode extracted from L_i and that from E_j). Additionally, $Freq(L^p_{i,\theta_n}, E^q_{j,\theta_m})$ provides a penalty if the relationship between L^p_{i,θ_n} and E^q_{j,θ_m} is simply a steady activity, i.e., L^p_{i,θ_n} and/or E^q_{j,θ_m} occur too often.

To mine useful information, clustering should be executed by varying the clustering parameters. Given a range for each clustering parameter, we obtained multiple pairs of clusters and then provided a ranking list of the cluster pairs based on their scores (described in Definition 3).

Definition 4 (Diff($L^{p,A}_{i,\theta_n}, E^{q,A}_{j,\theta_m}, L^{p,B}_{i,\theta_n}, E^{q,B}_{j,\theta_m}$)).

$$Diff(L^{p,A}_{i,\theta_n}, E^{q,A}_{j,\theta_m}, L^{p,B}_{i,\theta_n}, E^{q,B}_{j,\theta_m}) = \|\mathsf{Corr}(L^{p,A}_{i,\theta_n}, E^{q,A}_{j,\theta_m}) - \mathsf{Corr}(L^{p,B}_{i,\theta_n}, E^{q,B}_{j,\theta_m}).\|$$

$\mathsf{Diff}(L^{p,A}_{i,\theta_n}, E^{q,A}_{j,\theta_m}, L^{p,B}_{i,\theta_n}, E^{q,B}_{j,\theta_m})$ measures the difference between groups in the co-occurrence of behavioral modes. $\mathsf{Corr}(L^{p,A}_{i,\theta_n}, E^{q,A}_{j,\theta_m})$ and $\mathsf{Corr}(L^{p,B}_{i,\theta_n}, E^{q,B}_{j,\theta_m})$ are the correlation in group A and group B, respectively. $L^{p,A}_{i,\theta_n}$ ($E^{q,A}_{j,\theta_m}$) shows segments in the p-th (q-th) cluster obtained from individuals belonging to group A. The greater the value of $\mathsf{Diff}(L^{p,A}_{i,\theta_n}, E^{q,A}_{j,\theta_m}, L^{p,B}_{i,\theta_n}, E^{q,B}_{j,\theta_m})$, the greater the difference in the frequencies of the co-occurrence of behavioral modes between the two groups.

Algorithm

Given a set of feature time series, i.e., location time series (L_1, ..., L_n) and other sensor data time series (E_1, ..., E_m), and ranges of clustering parameters, we preformed clustering for each time series and parameter and computed the scores of all cluster pairs obtained from L_i and E_j. This required efficient clustering of the time series; however, achieving efficient clustering was not trivial, because naively enumerating all cluster pairs was infeasible if the number and length of the time series were large. Therefore, we proposed an efficient method that overcomes this challenge. To segment time series efficiently, we first employed symbolic Aggregate Approximation (SAX) [15], which is a symbolization method, to convert a series of numerical values into a series of symbols. After obtaining the result of SAX, we reduced the length of the time series of symbols by merging the same adjacent symbols. Additionally, we used spectral clustering [16], which is a graph-partitioning algorithm and constructs a similarity graph, where each segment associated with a symbol corresponds to a graph node in our case. We regarded segments with the same label as a vertex in the graph, which speeds up the clustering process. An overview of the method is described in the Algorithm 1. Given a set of time series $\mathcal{X}$, a range of SAX parameter α, i.e., $[\alpha_{min}, \alpha_{max}]$, and a range of cluster number β, i.e., $[\beta_{min}, \beta_{max}]$, we first obtained features learned by an autoencoder neural network and added them to $\mathcal{X}$ (line 1). For each time-series X_i in $\mathcal{X}$, we computed its mean, μ, and variance, σ^2, to enable SAX-based segmentation (line 5). Next, for each parameter $\alpha \in [\alpha_{min}, \alpha_{max}]$, we executed time-series segmentation with SAX (line 7), followed by spectral clustering for each $\beta \in [\beta_{min}, \beta_{max}]$ (lines 8–12) (i.e., a segmentation result using α could be reused when β is randomized by the algorithm). The scores of all cluster pairs were then computed based on Definition 3, and results diversification was performed to remove similar co-occurring behavioral modes. We then divided the data according to the type of group,

calculated correlations in each group, and ranked the correlated behavioral modes based on the difference in correlation between the two groups.

Note that we employed SAX in the segmentation procedure because SAX uses the breakpoints of distribution of data for symbolization; thus, we could reuse these breakpoints to speed up the segmentation process. In addition, we wanted to avoid grouping data points with the same symbol into different clusters, we employed a graph-based clustering method (i.e., spectral clustering). Each function used in Algorithm 1 (e.g., SAX) is explained below.

Algorithm 1: Proposed method

Input: $\mathcal{X}$: a set of time-series, α_{max}: maximum #symbols, α_{min}: minimum #symbols, β_{max}: maximum cluster number, β_{min}: minimum cluster number

```
/* Feature learning */
X ← X ∪ LSTM-AUTOENCODER(X)
/* Clustering time-series */
for ∀X_i ∈ X do
    μ, σ² ← COMPUTE-MEAN-VARIANCE(X_i)   ▷ *[r]μ is mean and σ² is variance
    for α = α_max to α_min do
        SAX(X_i, μ, σ², α)
        MERGE-SYMBOLS(·)
        SIMILARITY-GRAPH-CONSTRUCTION()
        LAPLACIAN-EIGENMAPS()
        for β = β_min to β_max do
            R ← R ∪ {k−means++()}   ▷ *[r]store the clustering-result
/* Diversification */
R ← DIVERSIFICATION(R)
/* Find the co-occurrence behavior mode that differ greatly in
   frequency between two groups */
for ∀R_i ∈ R do
    G_{i,1}, G_{i,2} ← DIVIDE(R_i)   ▷ *[r]divide result into two groups
    D ← D ∪ {DIFFERENCE(G_{i,1}, G_{i,2})}   ▷ *[r]calculate the difference between two groups and store the result
```

Feature extraction/learning

We first extracted manually designed features from the time series (i.e., locomotion features, such as speed and angular speed, were extracted from a trajectory time series.) (Fig. 2a). To automatically obtain additional features from the raw time series or the basic handcrafted features in $\mathcal{X}$, we employed a Long Short-Term Memory (LSTM) autoencoder that comprises LSTM cells [17, 18]. An autoencoder neural network (Fig. 3) was then trained to output values equivalent to the input values. The autoencoder first used the encoder to compress the input into a latent representation and then used the decoder to reconstruct the output from this representation. By limiting the number of hidden units to less than the number of input and output units, the network could obtain compressed representations of the input while preserving important information of the input that was necessary for its reconstruction (i.e., the output of the autoencoder). Because locomotion modes, such as cruising, included in an animal trajectory could be regarded as latent representations of movements, concepts similar to locomotion modes were expected to be obtained by feeding handcrafted features (i.e., features designed by researchers) into the autoencoder. The network used in this study comprised an input layer, a hidden LSTM layer using a rectified linear unit activation function, and an output layer. We used output time series for each hidden unit (unit in the hidden layer) as inputs to the next procedure. The feature learned by each hidden unit was called a "learned feature". The input of the encoder and the output of the decoder corresponded to handcrafted features

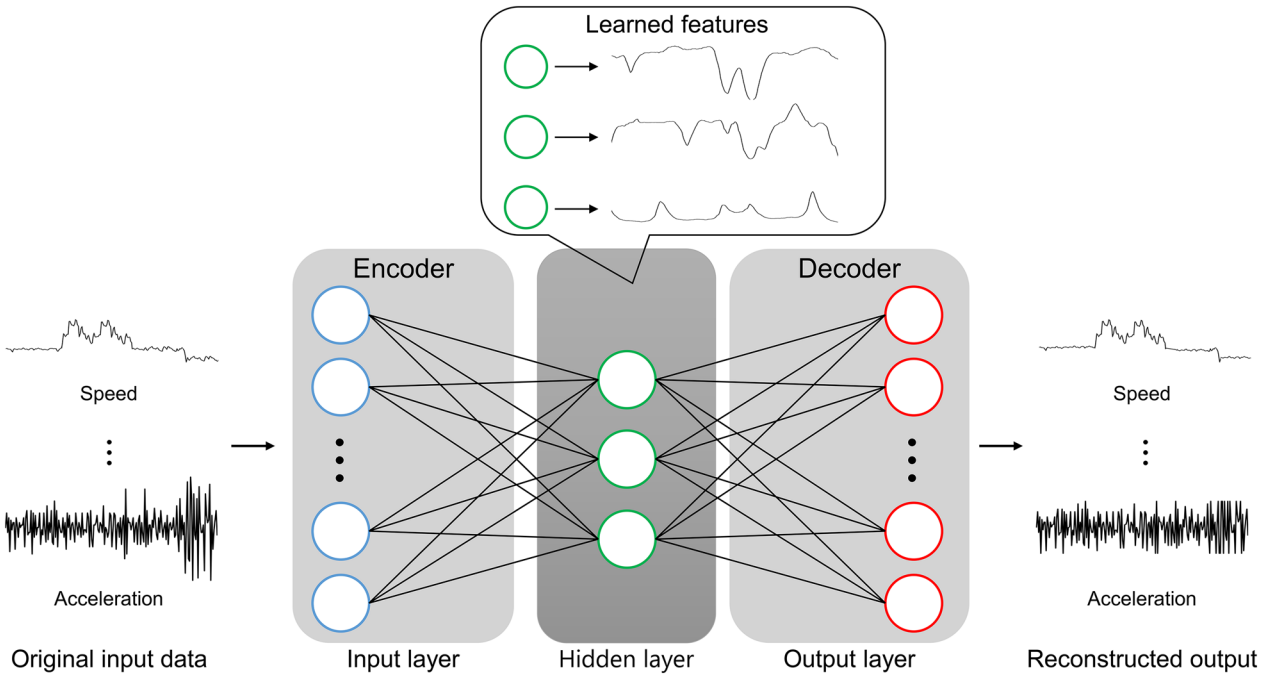

Fig. 3 Feature learning using an autoencoder. An autoencoder is a neural network that learns the latent representation of an input. The encoder converts the input into the latent representation, the hidden layer contains the latent representation of the input data, and the decoder reconstructs the original input from the latent representation. Each circle is regarded as a unit in the neural network. The number of units in the input layer and that of the output layer are identical to the dimension of the input data. The number of units in the hidden layer is less than that in the input layer. By limiting the number of hidden units, we can compress the input data. Because the compressed representation contains important information necessary to reconstruct the output, the compressed data are considered to represent latent states of an animal (i.e., behavioral modes)

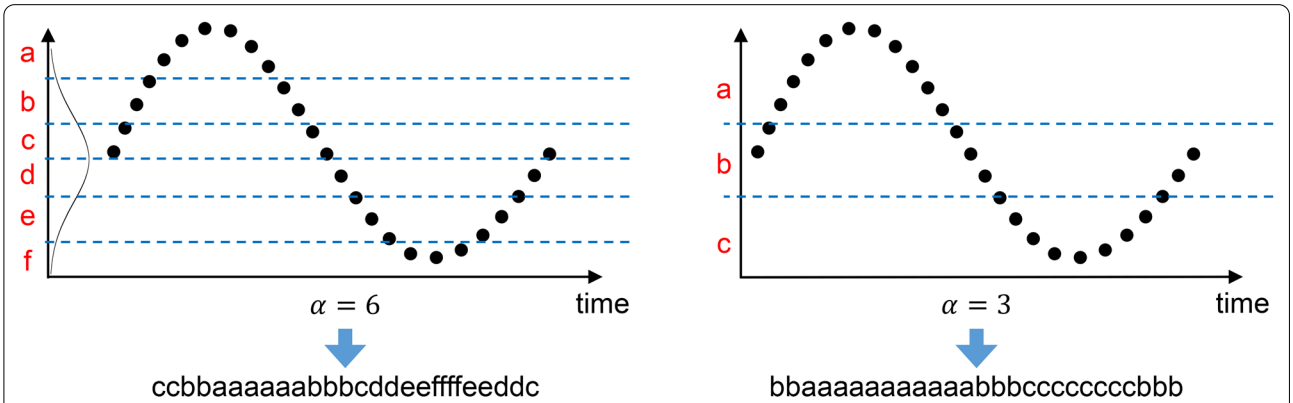

Fig. 4 Symbolization using SAX with different parameters ($\alpha = 6$ and $\alpha = 3$). The area under Gaussian distribution is divided into α equal-sized areas, with each area associated with a unique symbol. When a value in a time series falls into a divided area, the value is converted to a symbol corresponding to the space. Comparison of breakpoints between $\alpha = 6$ and $\alpha = 3$ shows that the second biggest breakpoint when $\alpha = 6$ is equal to the biggest breakpoint when $\alpha = 3$, the second biggest breakpoint when $\alpha = 6$ is equal to the biggest breakpoint when $\alpha = 3$. It means that the data converted to a and b when $\alpha = 6$ are converted to a when $\alpha = 3$; the data converted to c and d when $\alpha = 6$ are converted to b when $\alpha = 3$; the data converted to e and f when $\alpha = 6$ are converted to c when $\alpha = 3$. Therefore, we can easily obtain symbol series when $\alpha = 3$ by re-using that of $\alpha = 6$

in our study. We trained the network to minimize the mean absolute error between the input and output by employing backpropagation based on Adam [19]. We performed the feature learning procedure for each sensor modality. For example, we prepared an autoencoder for GPS data and features from the GPS data such as speed and acceleration were fed into the autoencoder.

Behavioral mode extraction

Behavioral mode extraction (Fig. 2b) included two phases: time-series segmentation and clustering.

Segmentation

Given a time-series X_i, we computed its mean, μ, and variance, σ^2, which were then used for all parameters $\in [\alpha_{min}, \alpha_{max}]$. We then segmented X_i using SAX, which assumes that normalized time-series data (based on μ and σ^2) $x \in X_i$ follows the Gaussian distribution and uses breakpoints (a sorted list of numbers) to divide the area under Gaussian curve to equal-sized areas. The data above the biggest breakpoint were mapped to the symbol `a` and the data smaller than the biggest breakpoint and larger than or equal to the second biggest breakpoint were mapped to the symbol `b` (Fig. 4). The symbols of the neighbor area were called *neighbor symbols*. The symbolic version of the time-series was `ccbbaa....eddc`, and the neighbor symbols of `c` are `b` and `d`. Note that as shown in Algorithm 1 (line 6), segmentation was executed in the descending order of α in order to reuse the previously obtained results. Specifically, when $\alpha = \alpha_i$, we reused the results of $\alpha = 2 \cdot \alpha_i$ to obtain the results of $\alpha = \alpha_i$. For example, if x was converted to `a` or `b` when $\alpha = 6$ in the case of Fig. 4, x was absolutely converted to `a` when $\alpha = 3$ (i.e., just converting `b` in the results of $\alpha = 6$ to `a` to obtain results of $\alpha = 3$).

Next, let $X_i(\mathtt{z})$ be the set of $x \in X_i$, where x was converted to `z`. We computed the average value of $X_i(\cdot)$ (i.e., avg $X_i(\cdot) = \frac{\sum x}{|X_i(\cdot)|}$) for each symbol, because this was one of the inputs to spectral clustering. Similar to symbolization, this averaging could utilize the previous results to save computation cost. In Fig. 4, we computed avg $X_i(a)$ when $\alpha = 3$ based on avg $X_i(a)$, avg $X_i(b)$, $|X_i(a)|$, and $|X_i(b)|$ was computed when $\alpha = 6$ (i.e., $\frac{|X_i(a)| \text{ avg } X_i(a) + |X_i(b)| \text{ avg } X_i(b)}{|X_i(a)| + |X_i(b)|}$). We then segmented X_i (line 7 in Algorithm 1). Each repeating symbol was coded once together (i.e., the segments of the time-series in Fig. 4 ($\alpha = 6$) was `c`, `b`, ..., `d`, and `c`).

Clustering

After X_i was segmented, we constructed a similarity graph defined, as follows.

Definition 5 (Similarity graph). Given a set of segments of X_i ($\{s_1, s_2, ...\}$), a similarity graph, G, constructed by the segment set is a weighted graph that has a node set and an edge set. Nodes correspond to the segments, and an edge is created between two nodes with the same symbol or neighbor symbols. Assume that two nodes, s_i and s_j, have an edge, and that their symbols are, respectively, `z` and `z'`. The weight of the edge w_{s_i,s_j}, which corresponds to the similarity between nodes, is defined below.

$$w_{s_i,s_j} = \frac{1}{|\text{avg } \mathrm{X_i(z)} - \text{avg } \mathrm{X_i(z')}|}. \tag{1}$$

We used this edge weight to reduce computation costs in the following procedure. It is worth noting that creating edges between all nodes would allow acquisition of an *odd* cluster, which contains some symbols but not their neighbor symbols. To avoid this, we considered only the same and neighbor symbols. After constructing G, we obtained the non-normalized Laplacian matrix of G and compute its eigenvectors. We then clustered nodes (i.e., divide G into k subgraphs G_1, G_2, ..., G_k by k-means++ [20] based on the eigenvectors) and obtained clustering result $X_{i,\alpha-\beta}$ (lines 9-11 from Algorithm 1). (The parameter k is $\beta \in [\beta_{min}, \beta_{max}]$ and $\beta < \alpha$.) This graph division aimed to minimize the summation of the weights of cut edges. Let V_j be the set of nodes of G_j. This graph-division problem was formalized to minimize

$$RatioCut(G_1, G_2, ..., G_k) = \sum_{i=1}^{k} \frac{Cut(V_i, V \backslash V_i)}{|V_i|}, \tag{2}$$

where

$$Cut(V_i, V_{i'}) = \sum_{s_j \in V_i, s_{j'} \in V_{i'}} w_{s_j,s_{j'}}, \tag{3}$$

$(s_j, s_{j'})$ was a removed edge, and V was a node set of G. These operations incurred a high computational cost if the graph size was large; therefore, we reduced the computational cost by graph-size reduction via merging nodes with the same symbol into a single node. This merged similarity graph was defined as follows.

Definition 5 (Merged similarity graph). The merged similarity graph of G, $\hat{G}$, includes nodes with different symbols and edges that were created between nodes with neighbor symbols. Assume that two nodes, $\hat{s}_i$ and $\hat{s}_j$, in $\hat{G}$ have an edge, and that their symbols are, respectively, `z` and `z'`. Let $V(\mathtt{z})$ be the set of nodes having symbol `z` of V. The weight of the edge $\hat{w}_{s_i,s_j}$ is

$$\hat{w}_{s_i,s_j} = \sum_{s \in V(\mathtt{z}), s' \in V(\mathtt{z'})} w_{s,s'}. \tag{4}$$

Definition 5 shows that the numbers of nodes and edges of $\hat{G}$ are α and $O(\alpha)$, respectively. Note that $\alpha \ll |V|$ and α are small integers in practice, suggesting the efficiency of graph-size reduction. Furthermore, we showed that this reduction in graph size did not sacrifice its correctness.

Theorem 1 (Correctness). *$\hat{G}$ provides the same clustering result as that with G.*

Proof First, the weights of edges between nodes with the same symbol are regarded as ∞ from Eq. (1); therefore, nodes with the same symbol are never assigned to different clusters when $\alpha >= \beta$, thereby allowing nodes with the same symbol to be merged. Moreover, based on this observation and Eqs. (3) and (4), we have

$$\begin{aligned} Cut(V(\mathrm{z}), V(\mathrm{z}')) &= \hat{w}_{s_i,s_j} \\ &= Cut(\{\hat{s}_i\}, \{\hat{s}_j\}). \end{aligned}$$

This suggests that the cutting results of G and $\hat{G}$ are the same. Therefore, Theorem 1 holds. □

Theorem 1 allowed use of $\hat{G}$ instead of G, enabling efficient execution of spectral clustering. Next, we introduce the time complexity of the clustering step.

Theorem 2 (Time complexity). *Given a time series X_i, clustering requires $O(\sum_\alpha(|X_i| + \alpha^3 + \sum_\beta \alpha\beta))$ time.*

Proof SAX and MERGE-SYMBOLS(·) requires $O(|X_i|)$ time. Similarity graph construction incurs $O(\alpha^2)$ time, whereas computing the eigenvectors requires $O(\alpha^3)$ time [21]. For a fixed iteration number, k-means++ requires $O(\alpha\beta)$ time [22]. Because k-means++ is executed $(\beta_{max} - \beta_{min} + 1)$ times, for a given α, we require $O(|X_i| + \alpha^3 + \sum_\beta \alpha\beta)$ time. Therefore, the time complexity of the clustering step is $O(\sum_\alpha(|X_i| + \alpha^3 + \sum_\beta \alpha\beta))$, which completes the proof. □

Because α and β were practically small integers, the method could rapidly obtain all cluster pairs.

Identifying co-occurrences in behavioral modes and diversification

We computed a ranked list of the pairs (co-occurrence of behavioral modes) based on their scores (Fig. 2d), followed by removal of similar co-occurrences of behavioral modes in the list. We first grouped co-occurrences of behavioral modes derived from the same features. For each group, we then computed the time correlation between each pair of behavioral modes. Given two behavioral modes, $(L^p_{i,\theta_a}, E^q_{j,\theta_b})$ and $(L^r_{i,\theta_c}, E^s_{j,\theta_d})$, we computed $\mathrm{Corr}(L^p_{i,\theta_a}, L^r_{i,\theta_c})$ and $\mathrm{Corr}(E^q_{j,\theta_b}, E^s_{j,\theta_d})$. When the two correlation values exceeded a threshold, the co-occurrence of a behavioral mode with a lower score was discarded. This was because some features obtained by the autoencoder can be strongly correlated with each other.

Identifying the co-occurrence of behavioral modes that differs between groups

After we obtained the ranked list of co-occurrences of behavioral modes, we calculated the difference in the correlation between groups in descending order (Fig. 2c). First, for each co-occurrence of behavioral modes, we divided the corresponding animal-locomotion data or learned features into two groups (e.g., groups A and B). Second, we calculated $\mathrm{Corr}(L^{p,A}_{i,\theta_n}, E^{q,A}_{j,\theta_m})$ and $\mathrm{Corr}(L^{p,B}_{i,\theta_n}, E^{q,B}_{j,\theta_m})$, respectively. Next, we computed the absolute value of the difference between the two groups and then obtained a ranked list of differences between males and females. Finally, the co-occurrence of behavioral modes that differs in frequency between the two groups was output.

Dataset

We evaluated our method using two time-series datasets: the 2D trajectories from GPS and water-depth sensor data, which are the most widely used data types [23]. The GPS and water-depth data were collected from sensor data loggers (Axy-Trek; Technosmart, Roma, Italy) attached to Streaked Shearwaters (*Calonectris leucomelas*) breeding around Awashima Island (38° 28' N, 139° 14' E; Niigata, Japan) in 2016, 2017, and 2018. Analysis of these sensor data using the proposed method was expected to reveal co-occurrences of behavioral modes related to the relationship between diving activity and locomotion mode. Refer to [24] for details concerning data collection. The sampling interval of the GPS was 1 min, and that for the depth sensor was 1.0 Hz. The number of trajectories was 79, the average duration of the trajectories was 329 h, and the average number of data

Table 1 Description of feature parameters. These parameter values are selected to capture short- and long-term behaviors of seabirds

Features	feature parameters
Moving average of speed, acceleration and angular speed	window size (min) = 5, 10, 15, 20, 25, and 30.
Moving variance of speed, acceleration and angular speed	window size (min) = 5, 10, 15, 20, 25, and 30.
FPT	radius (m) = 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000.
Displacement	window size (min) = 5, 10, 15, 20, 25, and 30.

points in one trajectory was 14,233. The dataset included information for 37 males (average trajectory duration: 334 h; average number of data points in one trajectory: 14,459) and 42 females (average trajectory duration: 325 h; average number of data points in one trajectory: 14,034). Because we could obtain a multi-dimensional time-series for each trajectory (79 time-series in total), we concatenated the 79 time-series to make one multi-dimensional time-series, and used this as input.

Experimental methodology

We prepared three types of time-series: 1) locomotion features from GPS, 2) the number of dives extracted from water-depth data, and 3) time stamps. We extracted locomotion features widely used in biology studies, such as speed [m/s], acceleration [m/s^2], angular speed [rad/s], moving average and variance of speed, acceleration and angular speed, first passage time (FPT) [13] calculated from GPS coordinates, distance to coastline, and displacement [12, 25], by varying their feature parameters (e.g., window size). Table 1 shows the description of the feature parameters. For example, we computed the moving average of speed with different windows size (5, 10, 15, 20, 25, 30), resulting in six time-series of the moving average of speed. Because the distance to coastline has no feature parameters, for example, a single time-series was computed. We also computed the displacement from the GPS data using sliding window with different windows size (5, 10, 15, 20, 25, 30), resulting in six time-series of displacement.

In total, 55 feature time-series data were extracted from each trajectory, from which we obtained 32 features learned by the autoencoder. We set α_{min} to 4, α_{max} to 20, β_{min} to 2, and β_{max} to 10, respectively. These parameters were determined empirically. We obtained a depth time series with sampling intervals of 1 min and 1 h by calculating the number of dives per minute and hour, respectively. Additionally, we used the series of time stamps, which were based on the hour of the day (e.g., 01:00, 02:00, etc, local time), as the third feature type, because we expected to obtain a correlation between the diving activities or locomotion features of the seabirds and the hour of the day. Note that, when we computed two feature time series, their sampling interval should be identical; therefore, we converted the depth data into the time series of the number of dives so that the sampling interval of the time series was identical to that of the GPS data (or the series of time stamps). α ranged from 4 to 20, and β ranged from 2 to 3. The following methods were used to evaluate the computational time of the proposed method. In this study, we focused on the computational time. In general, researchers iterate analysis processes many times by designing and adding new features according to the results of the previous run. Therefore, reducing the computational cost of an iteration is crucial.

- Proposed: the proposed method in this paper.
- w/o symbol: a variant of the proposed method that does not reuse previous SAX results.
- w/o reduce: a variant of the proposed method that does not use a merged similarity graph.

Note that all the above methods provided the same output. However, the computation times were different. Furthermore, we omitted cluster pairs not satisfying $\mathsf{Corr}(L^p_{i,\theta_n}, E^q_{j,\theta_m}) > \gamma$ in order to filter out behavioral modes with low correlation. Additionally, we set γ to 0.3. All methods were implemented in C++ and eigenvectors were calculated by CLAPACK 3.2.1. Moreover, we ran K-means++ and HMM on our data set instead of the combination of SAX and spectral clustering. As these methods do not have the segmentation procedure, we simply varied the number of clusters from 3 to 10. K-means++ was implemented in C++. HMM was implemented in Python by using hmmlearn 0.2.5. All experiments were conducted on a PC with a 3.4-GHz Intel Core i7 CPU with 16 GB RAM.

Results

Table 2 shows the running time of each method. The segmentation and clustering columns show the computation times of the segmentation and clustering procedures, respectively, indicating the efficiency of the proposed method in terms of the computational time compared to the other methods.

42 co-occurrences of behavioral modes were obtained from the method. The following are the co-occurrences of behavioral modes with the top three highest scores. Note that, as the goal of the study was to find co-occurrences of two behavioral modes, we focus only on these two behavioral modes and ignore the remaining behavioral modes.

Co-occurrence of behavioral modes with the best score: "A behavioral mode corresponding to small values of a learned feature by the 23rd hidden unit in the LSTM autoencoder" and "a behavioral mode of distance from the coastline (from 60 to 2200 m)" were highly correlated. The Corr of male seabirds and female seabirds were 0.358 and 0.615, respectively. The learned feature related to the locomotion speed, because the feature was highly correlated with the average moving speed (window size: 30 min; Pearson correlation: 0.902); therefore, the behavioral mode from the learned feature was translated as low speed (between 0.1 and 6 m/s). To determine the meaning of a behavioral mode (e.g., low-speed mode), we simply plotted the distribution of the identified feature and

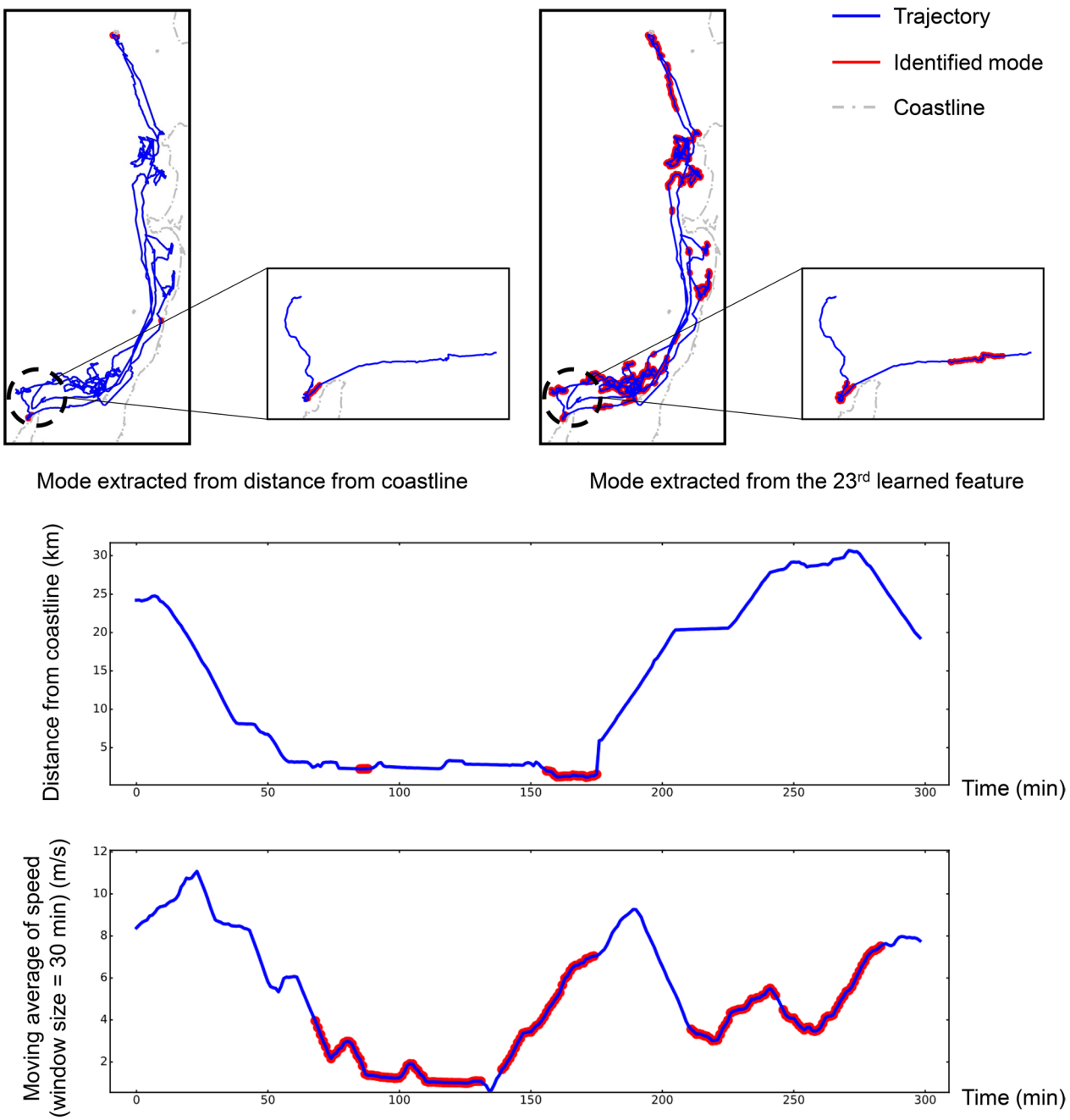

Fig. 5 Example of a co-occurrence of behavioral modes for female seabirds with the best score. The correlated behavioral modes are the behavioral mode of proximity to the coast (60–2200 m from the coast) from distance from coastline and the low speed mode from the 23rd learned feature. The upper maps show bird trajectories. The dashed gray lines show the coastlines, the blue line shows the trajectory, and the red segments represent identified behavioral modes. The lower graphs show time series of features, and the red segments show the identified behavioral modes

compared it to the distribution of data points belonging to the behavioral mode.

As above, the frequency of co-occurrence of the low speed mode and the behavioral mode of proximity to the coast for the female seabirds was higher than that for the male seabirds.

Figures 5 and 6 show the entire trajectories of a female and male seabird under the co-occurrence of behavioral modes with the best score. The red segments are the identified behavioral modes.

Co-occurrence of behavioral modes with the 2nd best score: "A behavioral mode of daytime (10:00–17:00 in

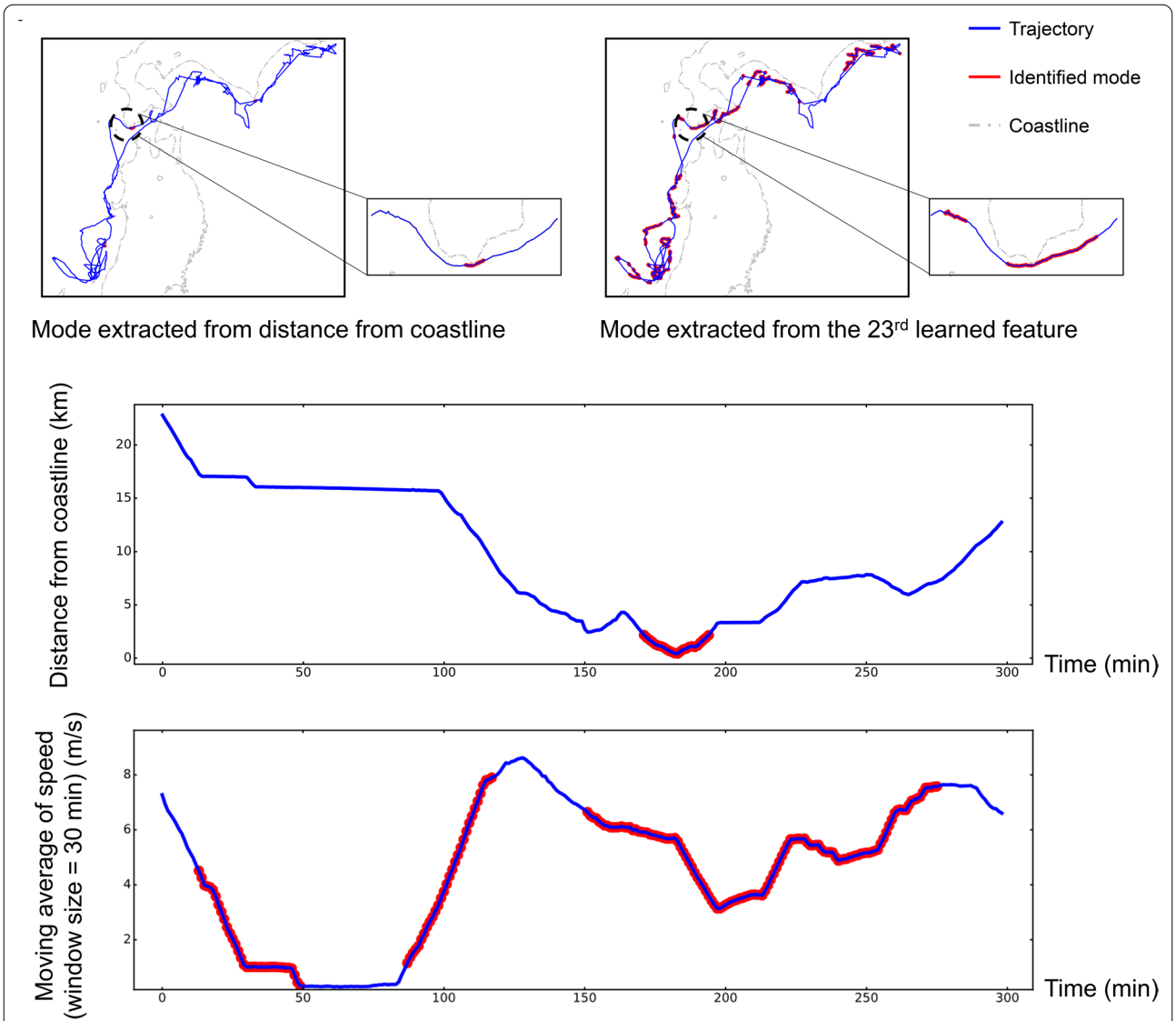

Fig. 6 Example of a co-occurrence of behavioral modes for male seabirds with the best score. The correlated behavioral modes are the behavioral mode of proximity to the coast (60–2200 m from the coast) from distance from coastline and the low speed mode from the 23rd learned feature. The upper maps show bird trajectories. The dashed gray lines show the coastlines, the blue line shows the trajectory, and the red segments represent identified behavioral modes. The lower graphs show time series of features, and the red segments show the identified behavioral modes

local time)" and "a behavioral mode of medium frequency of dives (between 10 and 14 times per hour)" were highly correlated. The Corr of male seabirds and female seabirds were 0.409 and 0.620, respectively.

As above, the frequency of co-occurrence of the daytime mode and the behavioral mode of medium frequency of dives for the female seabirds was higher than that for the male seabirds.

Additional file 1: Figure S1 and Additional file 2: Figure S2 show the entire trajectories of a female and male seabird under the co-occurrence of behavioral modes with the 2nd best score. The red segments are the identified behavioral modes.

Co-occurrence of behavioral modes with the 3rd best score: "A behavioral mode corresponding to moderate values of a learned feature by the 15th hidden unit in the LSTM autoencoder" and "a behavioral mode of multiple dives (i.e., behavioral mode corresponding to time when a seabird performed multiple dives per min)" were highly correlated. The Corr of male seabirds and female seabirds were 0.404 and 0.257, respectively. The learned feature was related to the locomotion speed, because the feature was highly correlated with the average moving speed

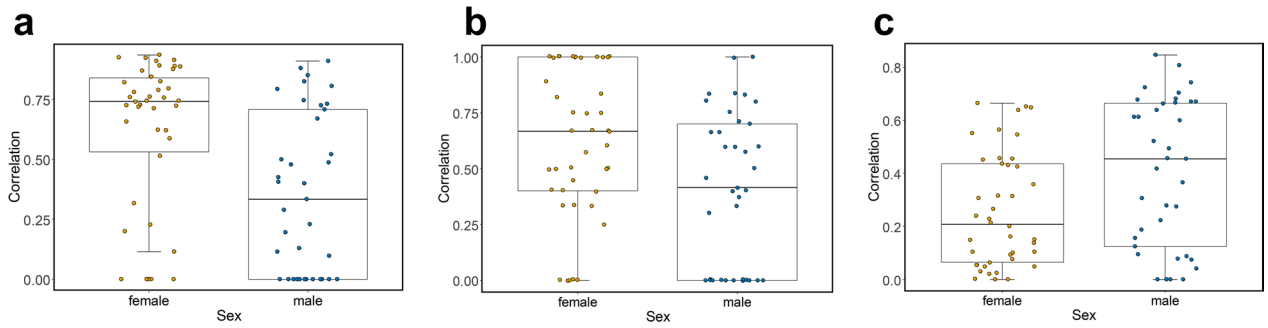

Fig. 7 Boxplot of Pearson correlation for each co-occurrence of behavioral modes. A significant difference between male and female seabirds was observed by the generalized linear mixed models with Gaussian distributions. Each data point corresponds to a trajectory. **a** Co-occurrence of behavioral modes with the best score ($t = -3.515$, $df = 77$, $p = 7.39 \times 10^{-4}$). **b** Co-occurrence of behavioral modes with the 2nd best score ($t = -2.788$, $df = 77$, $p = 6.68 \times 10^{-3}$). **c** Co-occurrence of behavioral modes with the 3rd best score ($t = 2.668$, $df = 77$, $p = 9.30 \times 10^{-3}$)

Table 2 Evaluation of the efficiency of seabird data: running time (s)

	Segmentation	Clustering	Total
Proposed	4.30	8.10	12.40
w/o symbol	5.98	8.16	14.14
w/o reduce	4.32	5.61×10^5	5.61×10^5
K-means++	–	–	173
HMM	–	–	1.28×10^5

(window size: 20 min; Pearson correlation: 0.914), and the behavioral mode from the learned feature was translated as a high-speed mode (> 5 m/s).

As above, the frequency of co-occurrence of the high-speed mode and the behavioral mode of multiple dives for the male seabirds was higher than that for the female seabirds.

Additional file 3: Figure S3 and Additional file 4: Figure S4 show the entire trajectories of a female and male seabird under the co-occurrence of behavioral modes with the 3rd best score. The red segments are the identified behavioral modes.

Figure 7 shows the boxplots of the correlation between females and males under the co-occurrences of behavioral modes with the top three scores. The co-occurrences of behavioral modes with the top three highest scores were analyzed by generalized linear mixed models (GLMMs) using Gaussian distributions. We treated Corr of each seabird as independent variable, sex as response variable and individuals as random factors in the models. We used R (version 3.6.0) package lme4 1.1.21 [26] for the linear models. The co-occurrence of behavioral modes with the best score differed significantly between males and females ($t = -3.515$, $df = 77$, $p = 7.39 \times 10^{-4}$). The co-occurrence of behavioral modes with the 2nd best score differed greatly between males and females ($t = -2.788$, $df = 77$, $p = 6.68 \times 10^{-3}$). The co-occurrence of behavioral modes with the 3rd best score differed between sex ($t = -2.228$, $df = 77$, $p = 9.30 \times 10^{-4}$).

Discussion

To the best of our knowledge, this is the first method that detects co-occurrences of behavioral modes that differ between two groups. Behavioral mode detection (time-series clustering) from multimodal data with unsupervised manner was problematic because meaningless sensor modalities affect the clustering results. In addition, the unsupervised methods required manual selection of parameters. Our method could deal with these problems by clustering time-series for each sensor modality and find the best combination of the modality and parameters efficiently.

As shown in Fig. 7, the proposed method could detect the difference in behavioral co-occurrences from multimodal time-series data recorded from wild seabirds. Additionally, the method substantially decreased computing time by applying graph-size reduction. As shown in Table 2, when comparing the proposed method with w/o symbolization, we confirm that re-using the previous SAX results reduced computation time related to the segmentation. Because $\alpha_{max} = 20$, the proposed method used the previous SAX results when $\alpha \leq 10$ (As α_{max} increases, we obtain benefits of re-using the previous SAX results). Moreover, comparison of the method without reducing the graph size indicated an increase in efficiency following the reduction in the clustering procedure. As shown in Table 2, the proposed method was fourfold faster following graph-size reduction. These results showed that reducing the graph size significantly affected the computational cost. The proposed method had superior performance in terms of computational time to that of either the K-means++ or HMM method,

indicating the high efficiency of the proposed method. Although the computational cost of K-means++ was much smaller than that of HMM, K-means++ could not take into account the temporal regularity of behavioral data.

In the experiment, we processed 87 (55 + 32) time series by changing two clustering parameters, resulting in 6,525 runs of the time-series segmentation and clustering process. Therefore, 21,284,550 combinations were investigated to find the best pair of time series and parameters. Our method could help this exhaustive analysis via fast time-series segmentation and clustering.

Investigation of the co-occurrence of behavioral modes with the best score indicated that the low-speed mode (from 0.1 to 6 m/s) and the behavioral mode of proximity to the coast (60–2200 m from the coast) were highly correlated, with females showing a higher correlation than the males. This co-occurrence of behavior mode indicates that the females fly at low speeds within 60 meters–2200 meters of the coast, which is more characteristic of females than males (Figs. 5 and 6). Because females are smaller than males, this suggests that they might prefer to stay close to the coast where the wind is not as strong [24].

The 2nd best score showed that the daytime mode (10:00–17:00 in local time) and the behavioral mode of medium frequency of dives (between 10 and 14 times per hour) were highly correlated, with females showing a higher correlation than males (Additional file 5: Figure S5 shows the ratio of multiple dives per hour for males and females). This indicates that females dive more frequently than males between 10:00 and 17:00 (Additional file 1: Figure S1 and Additional file 2: Figure S2). The differences in diurnal modes between males and females have been often observed in seabirds [27], which was interpreted as a result of sexual size dimorphism or sex-specific roles. In the case of Streaked Shearwaters, females conduct relatively short trips for chicks, suggesting that they dive soon after initiating foraging trips (i.e., in the morning). As above, a new insight related to diving of Streaked Shearwaters and timestamp was obtained by applying our method.

The 3rd best score showed that the high-speed mode (> 5 m/s) and the behavioral mode of multiple dives (i.e., corresponding to time when a seabird performed multiple dives per min) were highly correlated, with males showing a higher correlation than females. This indicates that the male seabirds performed multiple dives at higher speed and more frequently than female seabirds (Additional file 3: Figure S3 and Additional file 4: Figure S4). This might be associated with sexual differences in foraging behavior. As above, a new insight related to diving of Streaked Shearwaters was obtained by applying our method to locomotion data of Streaked Shearwaters.

The proposed method was designed to detect co-occurrence of behavioral modes that differs between two groups. When the groups are more than two, we can detect the difference using the standard deviation of the Corr in all groups.

Conclusions

The proposed method allowed acquisition of novel insights related to seabirds. The experimental results showed that the proposed method could find co-occurrences of behavioral modes that differed with groups and it took less time to run than did other machine learning methods. Because this method was designed to mine meaningful co-occurrences from any type of time-series, it can potentially be applied widely to discover relationships between locomotion modes and foraging activity [28], heart rate and group formation [29], and brain signals and behaviors [30]. Our future work will involve verification of our findings by measuring other modalities of sensor data (e.g., recording diving activities of the seabirds with video loggers). In addition, the proposed method could be extended to detect behavioral modes in more than two groups (e.g., juveniles, subadults, and adults).

Supplementary Information

The online version supplementary material available at https://doi.org/10.1186/s40317-021-00242-2.

Additional file 1: Figure S1. Example of a co-occurrence of behavioral modes for female seabirdswith the 2nd best score. The correlated behavioral modes are thebehavioral mode of frequent dives (between 10 and 14 times perhour) from number of dives per hour and the daytime mode(10:00–17:00 in local time) from timestamp. The upper maps showbird trajectories. The dashed gray lines in the map show thecoastlines, the blue line shows the trajectory, and the red segmentsrepresent identified behavioral modes. The lower graphs show timeseries of features, and the red segments show the identifiedbehavioral modes.

Additional file 2: Figure S2. Example of a co-occurrence of behavioral modes for male seabirdswith the 2nd best score. The correlated behavioral modes are thebehavioral mode of frequent dives (between 10 and 14 times perhour) from number of dives per hour and the daytime mode(10:00–17:00 in local time) from timestamp. The upper maps showbird trajectories. The dashed gray lines in the map show thecoastlines, the blue line shows the trajectory, and the red segmentsrepresent identified behavioral modes. The lower graphs show timeseries of features, and the red segments show the identifiedbehavioral modes.

Additional file 3: Figure S3. Example of a co-occurrence of behavioral modes for female seabirdswith the 3rd best score. The correlated behavioral modes are thebehavioral mode of multiple dives number of dives per minute andthe high-speed mode from the 15th learned feature. The upper mapsshow bird trajectories. The dashed gray lines in the map show thecoastlines, the blue line shows the trajectory, and the red segmentsrepresent identified behavioral modes. The lower graphs show timeseries of features, and the red segments show the identifiedbehavioral modes.

Additional file 4: Figure S4. Example of a co-occurrence of behavioral modes for male seabirdswith the 3rd best score. The correlated behavioral modes are thebehavioral mode of multiple dives number of dives per minute andthe high-speed mode from the 15th learned feature. The upper mapsshow bird trajectories. The dashed gray lines in the map show thecoastlines, the blue line shows the trajectory, and the red segmentsrepresent identified behavioral modes. The lower graphs show timeseries of features, and the red segments show the identifiedbehavioral modes.

Additional file 5: Figure S5. Ratio of multiple dives between male and female seabirds per hour.

Acknowledgements
We thank the Yoda Lab members for field assistance.

Authors' contributions
YT performed method design, software implementation, data analysis, and manuscript writing. TM conceived and directed the study and performed method design, and manuscript writing. JK performed data analysis and manuscript writing. DA performed method design and manuscript writing. TH performed data analysis and manuscript writing. SM performed data collection and data analysis. KY performed data collection and data analysis, and manuscript writing. All authors read and approved the final manuscript.

Funding
This work is supported by JSPS Kakenhi (JP16H06539 and JP16H06541).

Availability of data and materials
The datasets and codes used and/or analyzed during the current study are available in https://github.com/t-m-comp/diff_animal_behavior.

Declarations

Ethics approval and consent to participate
This study of the Streaked Shearwater was approved by the Animal Experimental Committee of Nagoya University and conducted with permits from the Ministry of the Environment, Japan.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Graduate School of Information Science and Technology Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan. [2]Graduate School of Environmental Studies, Nagoya University, Furou-cho, Chikusa, Nagoya, Aichi 464-8601, Japan.

Received: 4 September 2020 Accepted: 13 May 2021
Published online: 05 June 2021

References

1. Tinbergen N. Comparative studies of the behaviour of gulls (laridae): a progress report. Behaviour. 1959;15(1/2):1–70.
2. Martin P, Bateson PPG, Bateson P. Measuring behaviour: an introductory guide. 2nd ed. Cambridge: Cambridge University Press; 1993.
3. Kalan AK, Boesch C. Audience effects in chimpanzee food calls and their potential for recruiting others. Behav Ecol Sociobiol. 2015;69:1701–12.
4. Kays R, Crofoot MC, Jetz W, Wikelski M. Terrestrial animal tracking as an eye on life and planet. Science. 2015;348(6240):2478.
5. Studd EK, Landry-Cuerrier M, Menzies AK, Boutin S, McAdam AG, Lane JE, Humphries MM. Behavioral classification of low-frequency acceleration and temperature data from a free-ranging small mammal. Ecol Evol. 2019;9(1):619–30.
6. Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SS-H, Murthy M, Shaevitz JW. Fast animal pose estimation using deep neural networks. Nat Methods. 2019;16(1):117–25.
7. Teimouri M, Indahl UG, Sickel H, Tveite H. Deriving animal movement behaviors using movement parameters extracted from location data. ISPRS Int J Geo-Inform. 2018;7(2):78.
8. Bennison A, Bearhop S, Bodey TW, Votier SC, Grecian WJ, Wakefield ED, Hamer KC, Jessopp M. Search and foraging behaviors from movement data: a comparison of methods. Ecol Evol. 2018;8(1):13–24.
9. Adam T, Griffiths CA, Leos-Barajas V, Meese EN, Lowe CG, Blackwell PG, Righton D, Langrock R. Joint modelling of multi-scale animal movement data using hierarchical hidden markov models. Methods Ecol Evol. 2019;10(9):1536–50.
10. Branson K, Robie A, Perona P, Dickinson M. High-throughput ethomics in large groups of drosophila. Nat Methods. 2009;6:451–7.
11. Aghabozorgi S, Shirkhorshidi AS, Wah TY. Time-series clustering—a decade review. Inform Syst. 2015;53:16–38.
12. Bryce RM, Sprague KB. Revisiting detrended fluctuation analysis. Sci Rep 2012;2:315.
13. Fauchald P, Tveraa T. Using first-passage time in the analysis of area-restricted search and habitat selection. Ecology. 2003;84(2):282–8.
14. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Scienc. 1985.
15. Lin J, Keogh E, Wei L, Lonardi S. Experiencing sax: a novel symbolic representation of time series. Data Mining Knowl Discov. 2007;15(2):107–44.
16. Ng AY, Jordan MI, Weiss Y. On spectral clustering: analysis and an algorithm. In: International conference on neural information processing systems: natural and synthetic; 2001, pp. 849–56.
17. Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with lstm. Neural Comput. 2000;12(10):2451–71.
18. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313(5786):504–7.
19. Kingma D, Ba J. Adam: a method for stochastic optimization. In: International conference on learning representations; 2015.
20. Arthur D, Vassilvitskii S. k-means++: the advantages of careful seeding. In: ACM-SIAM symposium on discrete algorithms; 2007, pp. 1027–35.
21. Liu J, Wang C, Danilevsky M, Han J. Large-scale spectral clustering on graphs. In: International joint conference on artificial intelligence; 2013, pp. 1486–92.
22. Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge: Cambridge University Press; 2008.
23. Browning E, Bolton M, Owen E, Shoji A, Guilford T, Freeman R. Predicting animal behaviour using deep learning: Gps data alone accurately predict diving in seabirds. Methods Ecol Evol. 2018;9(3):681–92.
24. Matsumoto S, Yamamoto T, Yamamoto M, Zavalaga CB, Yoda K. Sex-related differences in the foraging movement of Streaked Shearwaters Calonectris leucomelas breeding on Awashima island in the sea of Japan. Ornithol Sci. 2017;16(1):23–32.
25. Helmuth JA, Burckhardt CJ, Koumoutsakos P, Greber UF, Sbalzarini IF. A novel supervised trajectory segmentation algorithm identifies distinct types of human adenovirus motion in host cells. J Struct Biol. 2007;159(3):347–58.
26. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw ARTICLEs. 2015;67(1):1–48.
27. Miller MGR, Silva FRO, Machovsky-Capuska GE, Congdon BC. Sexual segregation in tropical seabirds: drivers of sex-specific foraging in the brown booby sula leucogaster. J Ornithol. 2018;159(2):425–37.
28. Weimerskirch H, Pinaud D, Pawlowski F, Bost C. Does prey capture induce area-restricted search? A fine-scale study using GPS in a marine predator, the wandering albatross. Am Natural. 2007;170:734–43.
29. Weimerskirch H, Martin J, Clerquin Y, Alexandre P, Jiraskova S. Energy saving in flight formation. Nature. 2001;413(6857):697–8.
30. Rattenborg NC, Voirin B, Cruz SM, Tisdale R, Dell'Omo G, Lipp H-P, Wikelski M, Vyssotski AL. Evidence that birds sleep in mid-flight. Nat Commun. 2016;7:1–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.