Preliminary investigation of co-occurrence rule extraction for sex-specific behavior of Streaked Shearwater

Yiming Tian
Graduate School of Information Science and Technology, Osaka university

Takuya Maekawa
Graduate School of Information Science and Technology, Osaka university

Joseph Korpela
Graduate School of Information Science and Technology, Osaka university

Daichi Amagata
Graduate School of Information Science and Technology, Osaka university

Takahiro Hara
Graduate School of Information Science and Technology, Osaka university

Sakiko Matsumoto
Graduate School of Environmental Studies, Nagoya university

Ken Yoda
Graduate School of Environmental Studies, Nagoya university

Abstract

Recent rapid advancements in sensing and microcomputer technologies have led to an era of big behavioral data, which involves generating huge amounts of animal behavior data from sensor devices, such as small GPS loggers, that are attached to animals, e.g., birds and mammals. Owing to this, several big cross-disciplinary projects have focused on understanding animal behavior, pertaining to fields such as biology, neuroscience, and information science.

In addition to animal behavior data, environmental and biological data can be collected using sensors to elucidate the mutualistic relation between behavioral responses and changes in the environment. The behavioral data, e.g., animals' trajectories, generally comprise several modes, such as foraging and cruising. Similarly, the environmental and biological time-series data, e.g., heartbeat data, can also comprise several modes, such as relaxing and tense modes. By extracting frequent co-occurrence rules from these modes, their mutual relationship can be understood. For example, a frequent co-occurrence rule might state that the animals can concurrently exist in cruising and relaxing modes. One application of such rules is in the comparative analysis studies that are often conducted in biology, such as in studies of sexual dimorphism in seabirds.

This study proposes an efficient method for discovering frequent co-occurrence rules that differ greatly in frequency between two groups (e.g., male and female groups) that is based on time-series segmentation and clustering. Our method discovers frequent co-occurrence rules between pairs of modes by automatically selecting the clustering parameters and features that maximize the usefulness of the obtained rules. In addition, we employ a feature learning approach to automatically abstract high-level features from handcrafted features. To find a combination of clustering parameters and features that yields a meaningful rule, we iterate the time-series

clustering by varying these inputs, i.e., parameters and features, and then output rules with high usefulness scores. Since each parameter can have a variety of values and there are many kinds of features to be extracted, which results in an increased number of iterations of the segmentation and clustering process, this study proposes a frequent rule extraction method that efficiently computes each iteration of time-series segmentation and clustering. After having obtained the frequent co-occurrence rules, we can then find rules whose frequency differs greatly between the two groups.

Time-series segmentation and clustering methods have been used in previous work to extract modes from trajectories and time-series sensor data, with the time-series of feature values, e.g., time-series of movement speeds, extracted from the original time-series data and then segmented and clustered. However, in those previous studies, the features extracted from original time-series data along with the clustering parameters (e.g., thresholds) were objectively designed/selected by researchers based on their experience.

We demonstrated the effectiveness of our method using actual animal locomotion data collected from seabirds: GPS and water depth time-series data. For the water depth data, we output a time series representing the number of dives, while locomotion features were extracted from GPS data, such as speed, acceleration, and FPT. Additionally, we used the series of timestamps, which were based on the hour of the day (e.g., 1 a.m., 2 a.m., etc.), as another feature, since we wanted to find out if there is any correlation between the diving activities or locomotion features of the seabirds and the time periods. Here we introduce an example co-occurrence rule that was greatly different between male and female seabirds. The rule relates to the number of dives and the time. Specifically, the mode of frequent diving and the mode of daytime (9 a.m. to 3 p.m.) are highly correlated, indicating that the seabirds dive many times from 9 a.m. to 3 p.m., and there is also a great difference between males and females under this rule. Figure 1 shows histograms generated according to the correlation of the two modes, which show that between 9 a.m. and 3 p.m. the males dive more frequently than the females. In addition, our method revealed new biological findings related to these animals that had not been previously discovered by biologists, offering a substantial contribution to traditional data analysis on biological data.

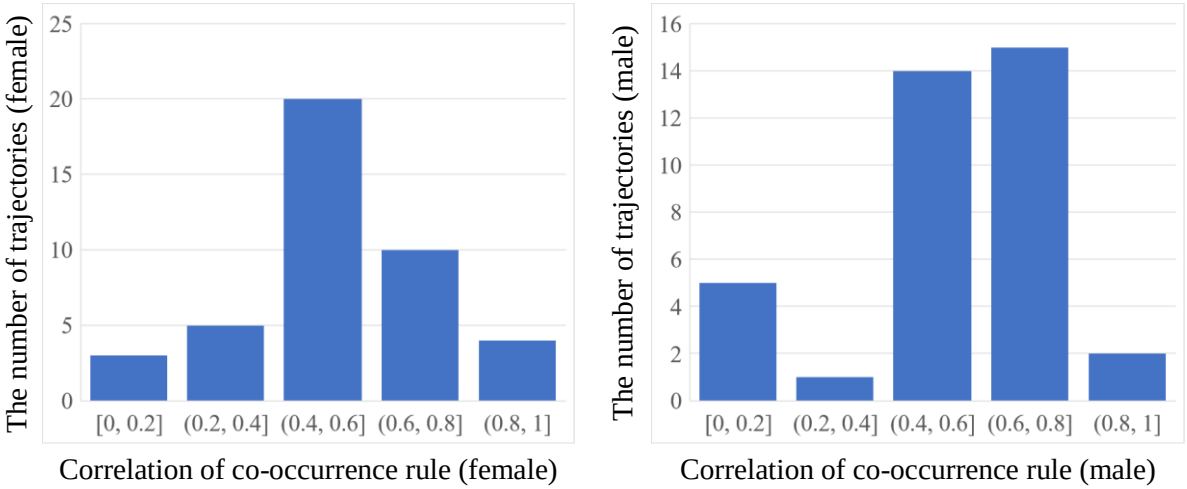

Figure 1. An example result where male and female seabirds differ greatly in a co-occurrence rule