Prague University of Economics and Business
Faculty of Informatics and Statistics

Automatic detection of life events in animal tracking data

MASTER THESIS

Study program: Knowledge and Web Technologies
Specialization: Quantitative Analysis
Field of study: studijní obor

Author: Karel Douda
Supervisor: RNDr. Ing. Petr Máša, Ph.D.

Prague, December 2022

Acknowledgements

-

Abstrakt

-

Klíčová slova

geospatial timeseries

JEL klasifikace

JEL1, JEL2, JEL3

Abstract

-

Keywords

geospatial timeseries, animal tracking

JEL classification

JEL1, JEL2, JEL3

Contents

List of Figures

List of Tables

List of abbreviations

BCC Blind Carbon Copy
CC Carbon Copy
CERT Computer Emergency Response Team
CSS Cascading Styleheets
DOI Digital Object Identifier
HTML Hypertext Markup Language
REST Representational State Transfer
SOAP Simple Object Access Protocol
URI Uniform Resource Identifier
URL Uniform Resource Locator
XML eXtended Markup Language

Introduction

Traditionally, the study of animal movements and life cycles has been a domain of great uncertainty, owing to varying levels of difficulty in observing the full range of animal behaviour in the wild. This has been especially difficult for ornithologists studying migratory birds. With the advent of small electronic circuitry, global telecommunications and positioning systems, it has become easier to acquire the required primary data for study. With the increasing sophisticatedness of GPS loggers, energy efficient solar panels, batteries and faster communication networks, another common problem has arisen, especially for larger studies. Modern animal movement loggers can hundreds, if not thousands, of positions per day, multiple times. It has become increasingly difficult for field experts to make sense of their data manually, while comprehensive data analysis providing useful results is not a trivial task.

Speedy detection of animal life events, such as mortality or nesting, is crucial for actionable instructions for animal conservation fieldwork experts. For example, speedy detection of animal death is very important for helping to establish the cause of the mortality event, since specific causes will be more difficult to establish later. Other life events

The goal of this diploma thesis is to provide animal conservation experts with reliable methods of filtering unreliable data, simplifying or clustering too complex datasets for interpretation and finally, detecting important life events from these filtered and clustered data. The structure of this diploma thesis roughly follows the CRISP-DM methodology, and the resulting models will be released on GitHub and integrated into the Anitra platform for animal conservation experts.

1. Problem context

1.1 Animal tracking

define animal tracking, brief history, problems, communication flow / time delays

1.2 Geospatial data

define geospatial data, precision problems, put it into context of animal tracking

1.3 Timeseries data

define timeseries, problems with existing methods for this problem (non-continuous sampling), possible solutions

1.4 Previous research

2. Data understanding

describe data sources, available columns, data ingestion

2.0.1 Provenance

2.0.2 Data description

0 datetime 224466 non-null object 1 gnss_latitude 224466 non-null object 2 gnss_longitude 224466 non-null object 3 gnss_altitude_m 224466 non-null object 4 gnss_speed_kmh 224466 non-null object 5 gnss_direction 224466 non-null object 6 battery_mv 224466 non-null object 7 battery_pct 224466 non-null object 8 solar_pct 224466 non-null object 9 temperature_c 224466 non-null object 10 baro_pressure_hpa 224466 non-null object 11 baro_altitude_m 224466 non-null object 12 roll 4 non-null object 13 pitch 4 non-null object 14 device_id 224466 non-null object 15 device_code 224466 non-null object 16 device_alias 4 non-null object 17 tracking_id 224453 non-null object 18 animal_id 224453 non-null object 19 animal_code 224453 non-null object 20 animal_name 4 non-null object 21 wrong_gps 224466 non-null object 22 note 4 non-null object 23 acc_raw_x 4 non-null object 24 acc_raw_y 4 non-null object 25 acc_raw_z 4 non-null object

gnss datetime, latitude, longitude, temperature

2.0.3 Data enrichment

elevation, temperatures, pressure

3. Data preparation

Data preparation is the phase concerned with constructing the final dataset, including record and attribute selection, data cleaning, construction of new attributes and transformation of data for modeling tools (Wirth et al., 2000). In the context of this work, several data preparation tasks were performed.

3.1 Adding new attributes

For further analysis requirements, several new attributes had to be added. None of the attributes are complex and are relatively fast to add in one pass through of the data.

Inter-position distance, the distance between two points in metres, is calculated using the Geodesic formula. inter-position distance (Geodesic formula)

Time difference is calculated by subtracting the two timestamps of the measurement time.

Average speed is calculated as the inter-position distance divided by time difference.

Bearing angle is calculated using the (which?) formula.

3.2 Outlier detection

Outlier is an observation that appears to deviate markedly from other members of the sample in which it occurs (Grubbs, 1969). Outliers can be of many types, but in the context of this work, measurement errors are the most concering type, since not all methods are resilient towards them.

Measurement errors are inherent to the problem area; small IoT devices are heavily energy-constrainted and animal movements are very time sensitive, therefore there might not be enough time for the device to detect the error or take another data sample.. Global positioning systems usually come with several meters of inprecision from the actual position of the receiver, but these inprecisions may become exaggerated in areas with poor reception or radio-reflective surfaces.

For example, earlier GNSS-based tracking systems used roughly over 40% of their energy on getting GPS positions and had an average precision of 15 meters in open-air conditions (Jain et al., 2008)

The resulting deviations may be in tens of metres to several kilometers between two relatively close positions in cities or dense forests, for example.

There are many types of errors with varying levels of difficulty for automatic detection; while all of the errors are obvious to an expert, making an algorithm for all of them proved difficult.

Speed-based position error detection

In the context of animal movements, a valid upper bound for movement speed can be specified with reasonable level of certainness. Most birds, apart from short periods of hunting, do not usually exceed the speeds of (?source?), and it can be reasonably assumed that the upper limit for speed should be (x). (consult with Dušan)

Angle-based position error detection

3.3 Attribute selection

data cleansing - normalisation, outlier detection (rules), interpolation, ordering, simplification?

3.4 Data labeling

Due to the intended results of the analysis process, precise and quickly interpretable result requirements, labeling of the data is a necessity. Therefore, a data labeling module was added to the Anitra animal tracking application for expert classification. Datasets will be enriched with data labels for each specific model. However, classic unsupervised learning algorithms such as k-means are not

3.5 Data splitting

For further sub-analysis, each dataset can be split into transfer and static states. Transfer state is assigned to the time the animal is moving between two specific locations, static state describes any other position. Transfer states can be further analysed to discover foraging or nest-building activity, and static states can be further analysed to discover mortality states, nesting states or night roosts.

K-means activity classification

In previous research, Schwager et al. demonstrated that the well-known K-means unsupervised classification algorithm can be used to classify tracking data into various categories, in their study they demonstrated a capability of classifying data into two clusters based on activity (or lack thereof). Using speed, horizontal head angle and vertical head angle, they were able to produce animal-independent uniform clusters for determining animal activity and inactivity states. Unfortunately for more general applications, head angles may be impossible to determine. Loggers may be placed on other parts of the animal that may make it impossible to determine these angles (such as on the back or on the leg), or may rotated in various different ways. This may create inconsistent clustering not only between different animals, but between different loggers.

(try this with other attributes)

4. Modeling

created models

reinforcement learning?

4.0.1 Stationary position clustering

kernel density estimation, cluster statistics

4.0.2 Cluster classification

4.0.3 Mortality detection

General mortality detection

Device specific mortality detection

5. Evaluation

???

6. Deployment

???

Závěr

Závěr je povinnou částí bakalářské/diplomové práce. Obsahuje shrnutí práce a vyjadřuje se k míře splnění cíle, který byl v práci stanoven, případně shrnuje odpovědi na otázky, které byly položeny v úvodu práce.

Závěr k diplomové práci musí být propracovanější – podrobněji to je uvedeno v Náležitostech diplomové práce v rámci Intranetu pro studenty FIS.

Závěr je vnímán jako kapitola (chapter), která začíná na samostatné stránce a která má název Závěr. Název Závěr se nečísluje. Samotný text závěru je členěn do odstavců.

Bibliography

GRUBBS, Frank E., 1969. Procedures for Detecting Outlying Observations in Samples. *Technometrics*. Vol. 11, no. 1, pp. 1–21. Available from DOI: 10.1080/00401706.1969.10490657.

JAIN, Vishwas Raj; BAGREE, Ravi; KUMAR, Aman; RANJAN, Prabhat, 2008. wildCENSE: GPS based animal tracking system. In: *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 617–622.

SCHWAGER, Mac; ANDERSON, Dean M; BUTLER, Zack; RUS, Daniela, 2007. Robust classification of animal tracking data. *Computers and Electronics in Agriculture*. Vol. 56, no. 1, pp. 46–59.

WIRTH, Rudiger; HIPP, Jochen, 2000. CRISP-DM: Towards a standard process model for data mining. In: *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*. Vol. 1, pp. 29–40.

Appendices

A. Formulář v plném znění

B. Zdrojové kódy výpočetních procedur