An Expert System for Discovering Biogeographically Interesting Locations from Animal Movement Data

IVANA NIŽETIĆ KOSOVIĆ¹, KREŠIMIR FERTALJ¹, JOSIP KUSAK²
¹Department of Applied Computing
Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, 10000 Zagreb
²Faculty of Veterinary Medicine, University of Zagreb,
Heinzelova 55, 10000 Zagreb
CROATIA
ivana.nizetic@fer.hr, kresimir.fertalj@fer.hr, kusak@vef.hr

Abstract: GPS technology enables collection of moving object’s positions remotely. Recent research on moving objects concerns analysis of their movement to increase the knowledge about their movement patterns. Discovery of biogeographically significant locations such as a den, rendezvous sites or kill-sites is very important in order to gain insight into animals’ behaviour, habitat selection and predator–prey interactions. Animals interact with their environment in a complex way; their movement is conditioned with underlying geographical space and semantics. Existing computerized methods for discovery of significant locations take only raw positions and time stamp into account, without necessary animal characteristics and pertaining geographical space. We propose an expert system to discover significant locations which enables inclusion of knowledge about both intrinsic and extrinsic properties of animals. Our expert rules are adaptable to different application domains, based on characteristics of animal and algorithm parameters. The results of this study will be useful to community members engaged in studies of wildlife, but we believe that our method could be applied not only to different kinds of animals but also to other classes of moving objects.

Key-Words: Moving object, wolf, Canis lupus, spatiotemporal data, GPS tracking, wildlife, expert system, clustering

1 Introduction
Animal radio tracking and monitoring has been done for more than 40 years to collect animal locations data and to increase knowledge about movement and habits of animals [1]. The tracking was and still is done by triangulation of VHF signal, observations of footprints and other signs like urine or faeces, by finding predators kill sites, animal sightings, etc. Recently, animals can be equipped with GPS collars. A GPS collar can be programmed to read its GPS position periodically (e.g. every 30 minutes or every 6 hours) [2]. Data collected during some period of time can be analysed in order to discover movement patterns, sites important to certain animal such as den, kill-site and so on, which could be useful to get better insight in animals’ habits [3].

Considerable research of moving objects was done in telecommunications where data are collected from mobile phone users to predict next access point [4], in ubiquitous computing to enhance the level of service by analysis and prediction of drivers’ intentions [5] and on moving objects in general [6]. However, methods and algorithms used in these applications cannot be fully applied for the analysis and prediction of the movement of objects such as wild animals, since they do not take into account habitat features and its use by certain animal species. Animals can be considered to be a special class of moving objects that are moving almost free in geographical space, therefore their movement is not constrained by some sort of network. Their movement is much more conditioned with underlying habitat (food, cover, vicinity of urban areas etc.) than the human movement [2, 7]. Their movement is much more irregular, outspread and less intelligible than human movement, which make modelling of space, time and movement of wild animals challenging and complex problem [2, 8].

In [9] authors suggest the model for enriching trajectories of people with semantic information such as touristic attractions. Unlike in human movement, where semantic information is usually known in advance, in the world of animals it has to be discovered first. In order to extract context information, we developed a module for discovery of sites with some specific meaning to certain species of wild animal. The module is a part of the system we have developed – the system for analysis and prediction of movement of wild animals that helps experts in getting insight into animal movements and habits.
2 Problem Formulation

By biogeographically significant locations we considered parts of the space (i.e. sites) with some specific meaning to certain animal or animal group. In case of wolves, significant locations could be den, rendezvous sites or kill-sites. Discovery of such sites is very important in order to gain insight into animals’ behaviour, habitat selection and predator–prey interactions. It was usually done by experts applying their prior knowledge about species, certain animal and geographical space used by observed animals [3] or by statistical models which take only raw positions and time stamp into account [10, 11]. Discovery of such sites will ease collection of field data and samples by helping researchers to find animal’s prey or enabling different kinds of analysis of change in kill sites or bedding sites through the seasons or years. It can even facilitate the process of prediction of animals’ next movement.

We propose an expert system to discover significant locations, which enables inclusion of knowledge about animal characteristics and pertaining geographical space. To the best of our knowledge, we were the first to use an expert knowledge system in the process of discovery of biogeographically significant locations. Our objective was to develop a method could be adaptable to different application areas, i.e. species of wild animals, based on characteristics of animal and algorithm parameters which can be defined and adjusted by user [12]. Our aim was to automate process as much as possible, but not to forget one of the main characteristics of decision support systems: it should support rather than automate decision making [13].

Data about moving objects, spatial and temporal data, expert knowledge, visualisation and algorithms are currently managed in partial or incomplete solutions which cannot fully satisfy wildlife researchers’ needs [8]. We intended to create geo-information software which enables the management of all data about moving objects in one place including various data concerning animals, expert knowledge and algorithms.

3 Discovering Biogeographically Interesting Locations

In order to facilitate the process of decision making about certain sites of wild animals, we developed a module for discovery of significant locations (see Fig. 1). The whole system consists of other modules such as module for representation of geographical data, module for location identification, movement prediction module etc. Module for location identification is of the great importance in the process of discovering of biogeographically significant locations as well.

According to the objective on the method to be adaptable to different application areas, database is meta-modelled as it was proposed in [12]. Entities such as species and its characteristics, individual animals and its characteristics, geospatial environment, time intervals and periods are stored in spatially-extended relational database.

Drools rule engine [14] is wrapped in a web service with interface to enable the process of discovery of significant locations. An end user application communicates with database and with rule engine via web service (see Fig. 1). User of the application is a field researcher of wild animals. We will refer to him or her in text as “expert”.

To perform discovery of significant locations, rule engine requires the following input data:

- **Information about locations** resulting from the clustering algorithm in module for location identification. Information about moving objects involved in analysis (identifiers from database) is provided as well.
- **Rules** which have to be written in a proper form (described in more detail in following sections).

After processing the rules, web service returns a list of candidate locations and fired rules. The candidate locations are then presented to the expert who makes the final decision.

In following sections we will describe the process of location identification and knowledge representation, essential for the performance of the described module.

3.1. Location identification

Location (region, area) is a compact part of space which consists of set of points (positions in certain coordinate system). To manipulate with locations rather than points is semantically reasonable approach, because two points on the map might have very close but different coordinates and still having the same meaning.

Considering that biogeographically significant locations manifest in historical data (collected GPS positions of animal) as sites with dense points, they can be extracted from data by clustering algorithms. Sites with dense points can be those visited by animal very often or those where animal stayed longer.

We used DBSCAN (Density Based Spatial Clustering of Applications with Noise) algorithm [15], because it discovers arbitrary number of clusters based on density of points and can recognize “noise”, i.e. points that don’t belong to any cluster. In our case, „noise“ is any GPS position of animal that is surrounded with only a few other animal positions, meaning it is probably not important place for it (not often visited place or the place at which an animal did not spend much time).
Unlike in [9], we used DBSCAN algorithm on data collected during some period of time, not only considering one trajectory.

Fig. 2 shows an example of clusters extracted from the movement data and represented by convex hulls of points in each cluster, which we consider as locations.

Other clustering algorithms can be used as well. By varying parameters of clustering algorithms, different results can be gained. The decision which locations are going to be sent to analysis is left to expert.

After selecting the locations which are going to be analysed, a list of locations has to be prepared in an appropriate XML-form. Following the root tag “Locations”, all location are listed with their properties and values. The list of points is listed as well, including identifiers of points and animal from the database, enabling additional information stored in the database to be reachable from the web service. Additional analysis can be performed if it is required to express expert knowledge in the rules.

In the following example is given XML representation for the list of locations, which is the result of position clustering for one female wolf movement during one-year period. A total of 1228 points are collected, in 6 hours intervals. Particular animal had identifier 23 in database. Locations were named “Loc_1”, “Loc_2” etc.

Each location was described by its following characteristics:

- **Number of points** – Loc_1 in the example consists of 121 from 1228 female wolf positions.
- **Points** – All 121 corresponding point identifiers from database. First point in the example has identifier 342. More animals can be observed at the same point, thus one point can be listed more than once with different animal id.
- **Seasonal time analysis** – points counted by wolf seasons. Seasons are defined by expert as intervals of the year. In the example, Loc_1 consists of 36 points recorded in spring, 9 in summer, 19 in winter and 57 in autumn from 121 points.
The XML list of locations will be converted by the web service to the list of location objects with corresponding values of their properties. Those objects will then be sent to the rule engine.

### 3.2. Knowledge representation

There are many reasons to separate rules from the rest of application[14, 16]. An expert can understand rules better than programming code and it is easier to maintain rules that are separated from the code. Rule systems provide an explanation facility allowing us to audit how and why a particular decision was made.

Expert usually knows the characteristics of the site that is important to the animal. Due to the space-time property of movement, it is hard to discover that site visually, by just looking at the geographical map and observations (points). For example, if we want to find the wolf den, we could search for the area where female wolf spent more time in the spring and summer [3]. Of course, many other characteristics can describe wolf den, based on the type of habitat and so on.

Our aim is to use rule engine to find the locations that satisfy conditions defined by expert from all the locations identified in prior step. The main idea is to enrich rules with additional information about animal and location. Table 1 lists some of the common properties, grouped by those related to animal, location and both animal and location. Expert may add and use custom properties as well.

<table>
<thead>
<tr>
<th>Properties related to animal</th>
<th>Properties related to location (Spatial)</th>
<th>Properties related to animal and location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social status, Reproductive status, Age, Gender</td>
<td>Habitat, Prey density, Area</td>
<td>Temporal, Number of positions, Density of points</td>
</tr>
</tbody>
</table>

One example of rule for finding the den is written below. In this example, the rule is first expressed in natural language, then in first order logic expression and finally in Drools syntax [14].

*If location is visited by female wolf mostly in spring and is very dense, it is a candidate for a den site.*

\[
\begin{align*}
\text{rule "Dense location"} \\
\text{when: } & \$l: \text{Location (NumberOfCoord>50)} \\
& \quad \$o: \text{MovingObject (class== "Wolf", gender == "Female")} \\
\text{then: ...} \\
\text{end} \\
\end{align*}
\]

\[
\begin{align*}
\text{rule "Spring location"} \\
\text{when: } & \$l: \text{Location()} \\
& \quad \text{eval(Season($l, "Spring")} \\
\text{then: ...} \\
\text{end} \\
\end{align*}
\]

function boolean Season(Location l, String season){...}
3.3. Decision making
The process of discovery of significant locations can begin when the expert knows how to write rules and how to perform sites identification. We assume that all data about animals, geospatial information, time intervals and periods (in this case: seasons) are in database. Steps performed by expert using our system are following:

1. Expert chooses moving object (animal) and period of year to select data which are going to be analyzed. Data are presented on the map.
2. Expert chooses parameters of clustering algorithm. Results (locations) are presented on the map.
3. Expert may repeat steps 1 and/or 2 or may continue to the next step.
4. Expert writes rules or loads previously defined rules.
5. Rules and list of locations are sent to rule engine which executes them and result is returned to the expert. Result shows list of candidates for requested type of sites with listed rules that certain candidate satisfies.
6. Expert may repeat steps 4 and 5.

After the last step, expert can choose which location(s) he or she is going to consider the most likely one(s).

The significant location candidates are presented to the expert in the following form:

| Location LocationName: FiredRuleName1, FiredRuleName2, ... |

For instance:

| Candidates: |
| Location Loc_1: Spring location |
| Location Loc_2: Spring location, Dense location |
| Location Loc_n: Dense location |

One of main characteristics of decision support systems is to improve the effectiveness of the decisions, not the efficiency with which decisions are being made, thus the final decision is on the expert to make [13]. The expert can interact with the system, repeating certain steps until he or she comes out with the satisfying results.

3.4. Model validation
We used the GPS data obtained by tracking of two wolves and run the proposed algorithms to determine den sites. Wolves were non-reproducing females, which both were members of two different packs during tracking time. One of them was tracked during one year and another during three consecutive years, with one recapture and replacing of the GPS collar. Positions were recorded every 6 hours. All positions collected during certain year for each animal were included in the analysis.

We discovered all four known dens (one of one female wolf and three of another – three years in a row). Three of four dens were discovered using historical data, i.e. comparing to sites were dens were found previously by the use of “traditional” methods. One den was found during the summer of 2010 by the use of proposed algorithms and confirmed in the field.

By varying cluster algorithm parameters we can obtain narrower or wider locations, but still identify the den site. Although, 6-hour interval may seem very coarse, results showed that it was dense enough for finding a den.

We plan to evaluate our method in discovering wolf kill-sites as well. We expect 6-hour rate to be too coarse, so 1-hour or 15-minutes rate would be more appropriate for finding kill-sites, especially for kill-sites where smaller prey was killed.

Regardless to the rate, the algorithms performed for sites discovery and analysis are the same.

4 Conclusion
We have developed an expert system which facilitates wildlife experts in finding biogeographically significant locations of animals based on collected GPS positions. Discovery of such sites could ease collection of animal signs in the field, like den sites, kill sites, day beds etc. The advantage of the approach is the opportunity to use expert knowledge and context in which animals are moving or not moving.

We have proposed the integrated geo-information software in which it is possible to manage all data about moving objects in one place, including various data, expert knowledge and algorithms.

The results of this study will be useful to community members engaged in studies of wildlife, but we believe the same generic model can be applied not only to different kinds of animals but also to other classes of moving objects.
References:


