Exploration through Enrichment: A Visual Analytics Approach for Animal Movement

David Spretke
University of Konstanz
spretke@dbvis.inf.uni-konstanz.de

Halldor Janetzko
University of Konstanz
janetzko@dbvis.inf.uni-konstanz.de

Florian Mansmann
University of Konstanz
mansmann@dbvis.inf.uni-konstanz.de

Peter Bak
IBM Research Haifa
Peter.Bak@il.ibm.com

Bart Kranstauber
Max Planck Institute for Ornithology
kranstauber@orn.mpg.de

Sarah Davidson
Max Planck Institute for Ornithology
davidson@orn.mpg.de

ABSTRACT

The analysis of trajectories has become an important field in geographic visualization, as cheap GPS sensors have become commonplace and, in many cases, valuable information can be derived either from the data themselves or their metadata if processed and visualized in the right way. However, showing the "right" information to highlight dependencies or correlations between different measurements remains a challenge, because the technical intricacies of applying a combination of automatic and visual analysis methods prevents the majority of domain experts from analyzing and exploring the full wealth of their movement data. This paper presents an exploration through enrichment approach, which enables iterative generation of metadata based on exploratory findings and is aimed at enabling domain experts to explore their data beyond traditional means.

1. INTRODUCTION

Animal tracking has become a powerful means to study the ecology of many kinds of animals. Technology improvements continuously increase the number of species that can be tracked, the temporal resolution and accuracy of location measurements, and the ability to use other sensors to record additional information about the animals and their surroundings. After the technical challenge of recording the positions of animals has been mastered, initiatives such as *Movebank*[1] or *Global Tagging of Pelagic Predators*[2] are working to solve standardization, metadata and data provenance issues to make such studies reproducible and comparable to each other. These initiatives are also offering an increasing number of linkages to external global environmental datasets (e.g., weather data).

However, the analysis interfaces widely available are still basic, and applying a combination of automatic and visual analysis steps to enrich and explore animal movement trajectories requires significant coding experience. Therefore, researchers' valuable analysis time is often spent with time-intensive data cleaning and coding activities that are needed to verify a few hypotheses, rather than with exploring the full wealth of the recorded data sets.

Visual analytics that combine automated and visual analyses can help to empower the animal tracking community and to foster new insight into the ecology and movement of tracked animals. In this paper we propose a unique integrated system that empowers its users to interactively explore all attributes of animal tracking data sets through iterative data enrichment. The visualization, interactive exploration and data enrichment features are the essential components that allow the user to link the raw data to hypotheses and provide a feedback loop to improve the data upon which new insight can be gained.

[1]http://movebank.org
[2]http://www.gtopp.org

ACM SIGSPATIAL GIS'11, November 1-4, 2011. Chicago, IL, USA

2. RELATED WORK

Since raw movement data are both very complex, as they represent rough approximations of complex activities, and at the same time semantically poor, it is necessary to develop appropriate analysis techniques. Andrienko et. al. [1] combine density-based clustering techniques on a sample of trajectories with a user-driven visual refining process to extract cluster representants. Afterwards a classifier is built to enable classification of the remaining trajectories.

A similar but fully automatic method for trajectory segmentation and sampling of moving objects is proposed by Panagiotakis et. al. [5]. As an alternative, a partition-and-group framework for clustering of trajectories was introduced by Lee et. al. [4], which enables the discovery of common sub-trajectories. Furthermore, the work of Lee et. al. [3] focuses on the classification of trajectories using a feature generation framework for trajectory data.

Dealing with movement data that have semantic context attached leads to the field of multi-dimensional data analysis, as the additional context data can be seen as multi-dimensional time series. The *Attribute Explorer* [6] or the *XmdvTool* [7] show the benefits of interactive *dynamic queries*. By an interactive feedback of the result of a query, dynamic queries allow the user to quickly visualize and understand the connections between dimensions.

We followed these ideas by allowing the user to interactively select time intervals or intervals of attribute values to highlight interesting parts of the data and combine it with capable data enrichment features on movement data.

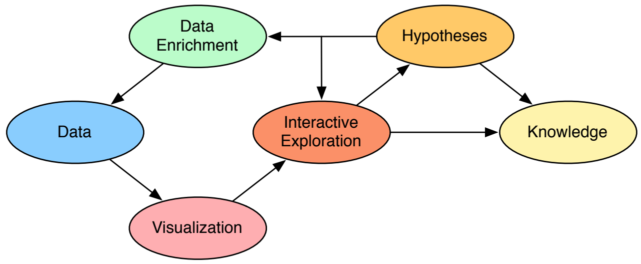

Figure 1: Exploration through Enrichment

3. THE EXPLORATION THROUGH ENRICHMENT APPROACH

The basic idea about our novel *exploration through enrichment* approach is depicted in Figure 1. The normal exploration workflow starts by visualizing a particular data set and then offering interaction capabilities for exploring the full wealth of the shown data. The user can instantaneously derive useful knowledge from this process or form hypotheses about the phenomena that she discovered in the data. In the next step, these new hypotheses can be rejected, confirmed or need to be checked in detail by starting the workflow over again. However, often the data needed to verify or explore the newly formed hypotheses is not available and must be created in a pain-staking external data enrichment process. In contrast to this, our tool integrates this vital enrichment procedure into the analysis interface. In summary, it is the interplay of the user with the system through interactive exploration and data enrichment that extends the analysis capabilities beyond traditional approaches.

4. ANIMAL ECOLOGY EXPLORER

This section introduces our visual analytics tool, the *Animal Ecology Explorer*, which is an implementation of the exploration through enrichment approach proposed in this paper. Figure 2 gives an overview of the visual interface of our tool, consisting a data loader window for selecting the animal tracking study; the data tracks (top left); the data attributes (below); several enrichment features, such as the Movebank weather interface, the segmentation window and the segment clustering window; map representations in the center and several configurable line charts and compressed time series graphs in the form of horizon graphs [2]. Brushing and linking is used to explore and relate behavioral or ecological phenomena. The following subsections describe the interactive exploration and data enrichment components of our system in detail.

4.1 Interactive Exploration

The Animal Ecology Explorer includes many components for interactive exploration of trajectory data and their associated metadata. In particular, these components are a geographic map interface combined with a track simplification method, several charts for metadata exploration and an interaction concept for relating the displayed data to each other.

Trajectory Visualization: In order to visualize geospatial datasets of moving animals, we naturally start with a map representation showing all selected trajectories. This map representation is based on the Java OpenStreetMap Framework; in addition to allowing layers to be drawn on top of the map, this framework offers geographic background tiles of satellite imagery (e.g., Microsoft Bing Aerial Map), elevation maps (e.g., OSM Cycle Map) or abstract geographic and political maps (e.g., TilesAtHome or Mapnik). This framework also provides basic interaction capabilities (e.g. zooming, panning, etc.).

A major challenge is that drawing trajectories of moving objects on a map often causes the problem of overplotting. Animals commonly stay in relatively small areas over extended periods of time, which results in regions with highly overlapping tracks. Furthermore, visualizing several tracks at the same time will result in even more visual clutter. In order to reduce this effect we apply a clustering algorithm to each single track to paint a simplified track (with less overplotting) on the map based on the current zoom level. In addition to showing the positions of the recorded data points, we can encode other variables in the data set by colorizing the tracks. The option to visualize one or two variables along the track is provided by colorizing the inner and outer portions of the track line. Using this approach we can view and compare two different variables along the track and view the relationships between these variables and location.

Metadata Exploration: Analyzing animal tracking data is very challenging as there are many variables that could explain the behavior and movement of an animal. We supply researchers with an overview time series visualization and detailed line charts for further investigations.

The compressed time series visualization in the form of a Horizon Graph [2], shown in the top part of Figure 2 detailing the altitude sensor measurement of gulls over a period of 1.5 years, uses a two-band color representation of selected attributes. If a value is above the average of the shown time series, it will be visualized by a blue hue, whereas a red hue is used for below-average values. Our representation is capable of comparing the same attributes for different animals or comparing different attributes for the same animal. This enables the user to see correlations between different attributes of the same individual and correlations between the same attribute of different individuals. While the absolute measurements are harder to estimate than in a line chart, the Horizon Graph is more scalable since it easily and clearly illustrates a lot of data without overplotting effects.

We implemented the line charts so that the user can view at maximum two animals and two attributes in the same chart. Color is used to differentiate between different animals, and dashed or solid lines denote the respective attributes. Color linkage thus enables cross-comparisons between the geographic extent of the individual animal's movement and the attributes displayed in the line charts.

Interaction: An essential part of our visual analytics tool is the interaction concept, which enables clear and efficient analysis of movement data. We connect all corresponding line charts and pixel visualizations with the geographic map representation using linking and brushing. By moving the mouse cursor within a line chart or pixel visualization, the user will automatically highlight the nearest point in the active chart and its corresponding points in all other shown visualizations, including the map. Furthermore, we provide dynamic filtering to select areas of interest in the map and highlight the corresponding regions in the line charts and pixel visualization. A very helpful technique for analyzing tracks is the dynamic filtering of attribute values and/or time/distance intervals, which can be done by marking regions within the plots to specify ranges of both time and attribute values for highlighting. Dragging the time slider with a defined range is a very helpful feature for biologists to interactively replay animal movement behavior. When attribute values are represented on the color attribute of the tracks, we omit color opaqueness, since combining the two color schemes together might lead to misinterpretations of opaque variants of the colors when matching it to the scale. In such a case, only the interactive movement of the slider reveals the directionality of movement.

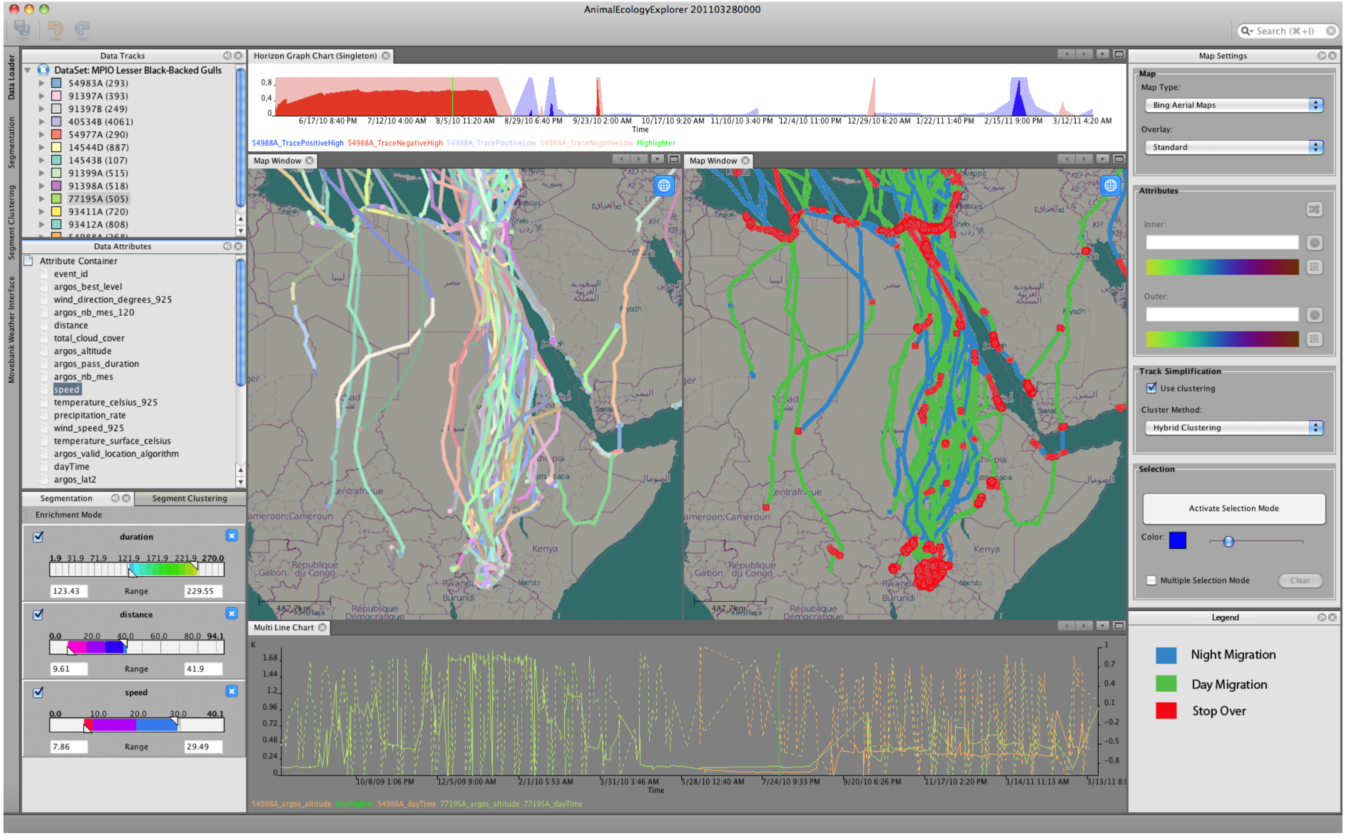

Figure 2: Visual interface of the Animal Ecology Explorer showing migrating gulls. On the left side the study and the tracked animals can be selected. A a horizon graph (top) and line chart (bottom) visualize attributes of individuals either over time or cumulated distance. The center depicts two zoom- and pannable map interfaces showing the segmentation of gulls' migration trajectories resulting in dozens of segments for every trajectory (center left) and clustering of segments with k-means (center right). Note the clearly distinguishable behavior for migration (purple) and for resting (azure with red border).

4.2 Data Enrichment

The unique feature of the Animal Ecology Explorer is the close integration of data enrichment with the interactive exploration components. However, the presented three components are just some of the many possible options to enrich trajectory data.

Basic Derived Attribute Calculations: Trajectories, as raw data, are usually time-stamped location informations. These are cleansed and preprocessed before being loaded into the Animal Ecology Explorer. While loading the tracks the tool derives additional attributes, such as *speed, distance, duration, time-related attributes, etc*. Once the trajectories are loaded, the tool creates a visual representation of their location on a cartographic map.

Data Access Modules for External Data: Access to external databases requires a lot of customization to meet the required data format and protocol of the data provider. In our case, we interface the Movebank API to enrich all coordinates of a set of trajectories with their associated weather conditions, such as temperature, wind speed and direction, geopotential height at different pressure levels, surface temperature, cloud coverage, or precipitation rate. For the end user, this is displayed as a system dialog with one checkbox for each weather parameter and hitting the "enrich" button results in new attributes in the data attributes window, which can then be used in the visualizations.

Formula Editor: Since our data model is mostly based on one record per GPS position, common movement parameters requiring more than one coordinate are automatically derived and then assigned to the coordinates. Our formula editor is a straight forward approach to derive attributes from the existing ones. To avoid complicated indexing syntax, we therefore restricted these methods to be either only applicable to single records or the whole dataset for aggregation operations (e.g., min, max and average). However, since basic calculations such as speed are already assigned to each GPS record, complex derived attributes can be calculated. Wind support for birds, for example, can be calculated using wind direction and strength as well as the birds flight direction.

Trajectory Segmentation and Annotation: The semantic annotation of trajectories is a fundamental task in understanding behavior from movement data. However, it is also a very challenging task, since GPS tracking devices reflect real world behavior, and are therefore very noisy, sometimes random, but mostly domain and context dependent. In order to achieve a semantic annotation that reflects the behavior of the objects under investigation, we suggest a tight integration of interactive visualization and automatic algorithms for information extraction. The role of the user is therefore to interact with the data and the algorithms through a visual interface to enhance the discovery process with his domain knowledge.

Trajectory Segmentation: Segmentation is conducted by setting threshold parameters as splitting criterion in the segmentation panel next to the data loader (see lower left part of Figure 2). The user can determine ranges of *speed, distance, and/or duration*, ac-

cording to which a trajectory is split into two or more segments. As an example, during bird migration, a resting period of at least 45 minutes, and a continuous flight distance of less than 2 km, indicates an interruption of the migration flow. This interruption can indicate sleeping places, feeding, etc.
The segmentation itself is carried out in a fully automatic manner by generating parameter driven queries to the spatial database. The setting of the parameters, however, is a fully manual task, carried out by the user. Every change of a setting results in an immediate response from the database, after which the data is re-rendered on the map. Colors are used to distinguish between consecutive segments by iterating through a set of qualitative colors.

Data Clustering: In many cases the definition of behavior is highly complex and based on multidimensional attributes. For such cases, the Animal Navigator provides a highly enhanced clustering feature. Clustering can be conducted by selecting one of dozens of standard algorithms and setting its specific parameters. The results of the clustering are immediately shown with additional statistical and quality information of the algorithmic performance. The user can iteratively optimize the parameter settings of the algorithm until the results are satisfactory. As a result of clustering, segments of trajectories are assigned to a single behavior type. Each behavior is described by a set of attributes and can now be semantically annotated.

5. EVALUATION

In this section, we describe a real-world case study about animal tracking that shows how our proposed tool is used by the domain experts co-authoring this paper. In this case study our investigation focuses on the migration behavior of gulls. A large-scale study was conducted to record the movement patterns of gulls during migration in the winter 2009 and 2010. There were 63 gulls (lesser black-backed gull, *Larus fuscus*) equipped with GPS receivers, recording on average 800 location signals during one migration period. These gulls breed in Europe and migrate annually from northern Europe to central Africa. During migration these birds can fly tremendous distances. Our investigation aimed to locate regions of resting places during their migration and try to find temporal differences in their migration patterns. The results of our investigation are shown in Figure 2. The segmentation window for setting parameters based on the loaded tracks and attributes is shown on the lower left corner. Results of the segmentation are displayed in the left map, and the right map shows the final results of the clustering and annotation. We first segmented the trajectories by setting thresholds for speed (< 4 km/h), distance (< 1 km) or duration (< 60 minutes). This resulted in 1,900 segments for the 63 trajectories. We then clustered these segments using a k-means algorithm (k=3). The clustering was conducted on the *average-speed* and *day-time* attributes of the data. For clustering we rely on the methods available in the WEKA[3] toolkit. There is a wide range of possible algorithms (e.g., DBScan) to match domain-specific requirements.
This segmentation resulted in three different clusters, which were annotated after investigating the cluster properties: 1. *Migration during day time*: described long distance flights with consistent speed, occurring during day time. 2. *Migration during the night*: described long distance flights with consistent speed, occurring during the night. In this cluster, segments had at least one location with a time-stamp at night. 3. *No migration segments*: described as short distance flights with large speed difference of consecutive segments or no-movement periods (resting) during the time period. Animals spend most of the year in this movement mode as they forage, rest and breed in a relatively restricted range.
These results illustrate a visual analytics approach to the semantic annotation of trajectories. Using this approach, domain experts can efficiently describe behavioral differences as movement patterns.

[3] http://www.cs.waikato.ac.nz/ml/weka/

6. CONCLUSIONS

In this paper we described a novel exploration through enrichment approach for trajectory analysis, which tightly integrates data enrichment features into the data analysis and exploration workflow. To demonstrate the applicability of this approach we presented the *Animal Ecology Explorer*, which is an interactive tool for confirmatory and exploratory analysis and enrichment of movement data. While each of its visual and analytical components is not novel in itself, our contribution is the tight integration of these components in a single analysis tool to not only help researchers to focus on data analysis rather than laborious coding, but also to empower them to make deep insights while exploring enriched movement data.

7. ACKNOWLEDGMENT

We thank the Movebank team Kamran Safi, Roland Kays and Martin Wikelski for the many rounds of feedback and for sharing their tracking data. This work was funded by the German Research Foundation (DFG) within the projects "Visual Spatiotemporal Pattern Analysis of Movement and Event Data" (ViaMod) and "MoveBank Virtual Research Environment" (MoveVRE).

8. ADDITIONAL AUTHORS

Additional authors: Manuel Mueller (University of Konstanz, `Manuel.Mueller@uni-konstanz.de`)

9. REFERENCES

[1] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive visual clustering of large collections of trajectories. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 3–10. IEEE, 2009.

[2] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1303–1312. ACM, 2009.

[3] J. Lee, J. Han, and X. Li. Trajectory Outlier Detection: A Partition-and-Detect Framework. In *IEEE 24th International Conference on Data Engineering*, pages 140–149. IEEE, 2008.

[4] J. Lee, J. Han, and K. Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 593–604. ACM, 2007.

[5] C. Panagiotakis, N. Pelekis, I. Kopanakis, E. Ramasso, and Y. Theodoridis. Segmentation and Sampling of Moving Object Trajectories based on Representativeness. *IEEE Transactions on Knowledge and Data Engineering*, 2011.

[6] R. Spence and L. Tweedie. The Attribute Explorer: information synthesis via exploration. *Interacting with Computers*, 11(2):137–146, 1998.

[7] M. O. Ward. Xmdvtool: integrating multiple methods for visualizing multivariate data. In *Proceedings of the conference on Visualization '94*, VIS '94, pages 326–333, Los Alamitos, CA, USA, 1994. IEEE Computer Society Press.