MigrO: a plug-in for the analysis of individual mobility behavior based on the stay region model

Maria Luisa Damiani
Dept. Computer Science,
University of Milan, Italy
damiani@di.unimi.it

Hamza Issa
Dept. Computer Science,
University of Milan, Italy
issa@gmail.com

Giuseppe Fotino
Dept. Computer Science
University of Milan, Italy
giuseppe.fotino@studenti.unimi.it

Fatima Hachem
Dept. Computer Science
University of Milan, Italy
hachem@di.unimi.it

Nathan Ranc
Dept. Organismic and
Evolutionary Biology Harvard
University, USA &
Dept. Biodiversity and
Molecular Ecology Fondazione
Edmund Mach, Italy
nathan.ranc@gmail.com

Francesca Cagnacci
Dept. Organismic and
Evolutionary Biology Harvard
University, USA &
Dept. Biodiversity and
Molecular Ecology Fondazione
Edmund Mach, Italy
francesca.cagnacci@fmach.it

ABSTRACT

We present MigrO, a clustering environment for the extraction of individual mobility patterns from GPS trajectories, relying on the notion of *stay region* [1]. A stay region is an 'attractive' area where the moving object resides for a period, possibly experiencing arbitrarily long periods of absence, before moving to a more attractive stay region. The core component is the SeqScan algorithm for the extraction of temporally ordered sequences of stay regions grounded on the notion of *presence*. An additional set of functionalities support trajectory pre-processing and clustering evaluation. MigrO is developed as plug-in for the open-source QuantumGIS system thus can exploit the rich set of functionalities of the hosting system, offering a formidable platform for the analysis of the mobility behavior. In this demonstration, we present MigrO at work in two case studies, both from the domain of animal ecology, illustrating two kinds of behavior, the migratory behavior and the exploratory behavior of two species of animals.

Categories and Subject Descriptors

H.2.8 [**Database management**]: Database applications—*Spatial databases and GIS*

General Terms

Design, algorithms

SIGSPATIAL'15 November 03-06, 2015, Bellevue, WA, USA

http://dx.doi.org/10.1145/2820783.2820794 ...$15.00.

Keywords

clustering, trajectories, behavior analysis, animal ecology

1. INTRODUCTION

The spatial trajectories that, nowadays, can be collected using modern location tracking technologies contain valuable information on the behavior of moving objects. Currently, a large body of research focuses on the extraction of *mechanistic* mobility patterns, namely patterns that have a straightforward, domain-independent formulation [2]. For example, the *flock* pattern is defined as a set of n moving objects which move in close proximity for a time interval k [2]. The problem with the mechanistic view is that the patterns are typically the result of an oversimplified interpretation of the real movement. Therefore, when deployed in real applications, they risk not to fit. In [1] we present a different, bottom-up approach. The idea is to start from data observation and domain knowledge to first develop an abstract model of the mobility pattern of concern, next translate such a model into an operational system. In particular, we propose an abstract model of the individual movement relying on the notion of stay region. A stay region is an 'attractive' area where the moving object resides for a period experiencing arbitrarily long periods of absence, before moving to a more attractive stay region. Indeed, the notion of stay region can be used for modeling diverse types of behaviors. For example, in the field of animal ecology, seasonal migrations can be seen as transitions from one stay region (the seasonal home-range), where the animal resides for most of the time during the seasonal period, to another seasonal home-range; conversely, the animals with a nomadic or exploratory behavior experience short period of residence in many different stay regions. In this sense, following the patterns taxonomy in [3], the stay region sequence represents a *generic* pattern.

To facilitate the study of the potentially wide spectrum of behaviors that can benefit from the notion of stay region, we have developed the MigrO system. The core component is SeqScan, the spatio-temporal clustering algorithm defined in [1] and based on DBscan [4] enabling the extraction of stay region sequences from input spatial trajectories. Unlike existing approaches, the technique does not impose any constraint on the amount of noise, while it assures that the staying periods are temporally disjoint. In addition, MigrO provides a set of tools supporting trajectory pre-processing and clustering evaluation, the latter through the analysis of a time-dependent index called *stationarity index*. MigrO is fully integrated into the popular Quantum GIS (QGIS) platform, thus can leverage the rich set of functionalities of the hosting system. A subset of the functionalities available in the current version has been employed for the clustering analysis reported in [5]. In this demonstration we present MigrO at work in two usage scenarios: a) analysis of the seasonal migratory behavior of a group of red deer (*Cervus elaphus*) located in the Bavarian region; b) analysis of the exploratory behavior of a group of roe deer (*Capreolus capreolus*) deliberately translocated from the native region into a territory in Southern Italy for conservation purposes. The rest of the paper is organized as follows: Section 2 overviews the mobility behavior model; Section 3 the MigrO architecture; the demo is outlined in Section 4 followed by conclusive considerations.

2. THE MOBILITY BEHAVIOR MODEL

We begin overviewing the key concepts underlying the system, next how these concepts have been translated into an operational system.

2.1 Abstract model

Figure 1 sketches the abstract, conceptual model of the individual mobility behavior built on the notion of stay region.

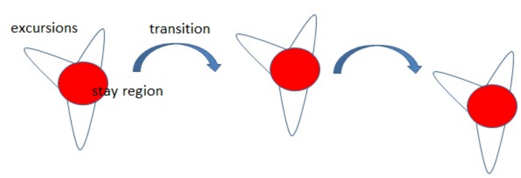


Figure 1: Abstract model: the object resident in a stay region (red circle) can perform excursions of arbitrary duration outside the region before moving to a different stay region.

Intuitively, a stay region is an area that for some reason, that is beyond our goal to investigate, results to be particularly 'attractive' for the individual in a given period of time. An attractive region implies that the object either stays within the region or, should it move outside for one or more *excursion*, the object likely returns back after an arbitrarily long time. Importantly, a stay region remains attractive until a more recent stay region is found. Thus, at any instant of the observation period, at most one attractive region exists. In this sense the concept of stay region resembles the more familiar notion of residence. Note that the notion of stay region is different from that of 'stop' or 'stay point'. In a stop the object is supposed to remain inside the region for the whole time before leaving definitely that area, while in a stay region the object can perform arbitrary long excursions. The goal of the analysis is to extract from a GPS trajectory the sequence of temporally disjoint stay regions at the desired temporal granularity. The result can contain one or multiple stay regions (and thus *transitions* from one region to the next) or even no region if the individual is stationary.

2.2 Operational model

Stay region. Given a trajectory consisting of a sequence of timestamped points $T = [(x_1, y_1, t_1), .., (x_n, y_n, t_n)]$, a stay region is represented by a sub-sequence of the input trajectory which may contain 'holes', i.e. missing points denoting periods of absence. A stay region has a duration which is simply defined by the time difference between the last and the first point of the sequence. The key features of the model can be summarized as follows: a) a stay region is defined as a DBscan cluster, thus consists of core points and border points, with respect to the input density parameters n and ϵ (i.e. it is a *dense region*); b) The dense region satisfies a temporal condition expressed in terms of 'presence' (δ is the presence threshold). Different from the concept of duration, the presence at time t in stay region c is an estimate of the cumulative time spent within c until t, excluding thus the periods of absence; c) The clustering algorithm processes multiple dense regions simultaneously. Dense regions are created, expanded, merged. The first dense region that satisfies the minimum presence requirement and which only contains points temporally following the active stay region becomes the new active stay region; the remaining points are added to noise. This processing strategy ensures that a stay region remains active until a more recent one is found. For the details we refer the reader to [1].

Excursions and transitions. The points of the trajectory that do not fall in any stay region do not constitute an undifferentiated noise. Rather the noisy points falling in the temporal extent of a stay region c represent the excursions performed by the moving object when resident in c while the noisy points falling in between the temporal extents of two consecutive stay regions, c and c', are classified as transitions [5].

Mobility indexes. The model comprises a few indexes which summarize relevant aspects of the object's behavior. The first is called *stationary index* [5]. It quantifies the degree of mobility within a stay region c_i. Specifically, the stationarity index $Q_{PD}(c_i)$ is defined as ratio of the cumulative presence $P(c_i)$ and the duration of the stay region $D(c_i)$. By extension, the average stationarity index Q_{avg} is: $\frac{1}{n}\sum_{i\in[1,n]}\frac{P(c_i)}{D(c_i)}$ where n is the number of stay regions extracted from the input trajectory. A low value of the stationarity index indicates relatively long periods of absence.

A different and novel index is called *commuting index*. It quantifies the degree of commuting between two stay regions, the *origin* and the *target*, in other words whether and to what extent, an object residing in one region visits the target region (either past or future). Specifically the index $Q_{CO}(c_i, c_j)$ is defined as ratio of the subset of excursions points in c_i that can be added to the target region c_j (i.e. are elements of the dense area) and the excursion points from c_i. A high value of the index indicates frequent commuting.

3. THE MIGRO ARCHITECTURE

The layered architecture is shown in Figure 2. MigrO

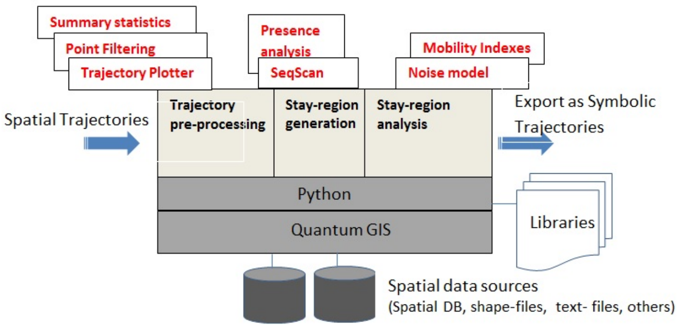


Figure 2: MigrO architecture

is developed in Python on top of QGIS. The system handles spatial trajectories represented by sequences of timestamped points, and returns sequences of stay regions that can take the form of labeled points or, in alternative, of symbolic trajectories [6]. The analytical process consists of three phases: trajectory pre-processing, spatio-temporal clustering for stay region generation, and stay region analysis for the assessment of the clustering outcome. Each phase is supported by a number of tools, briefly presented below.

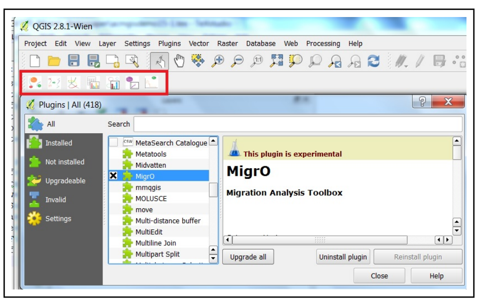


Figure 3: MigrO plug-in in QGIS V2.8.1: the red rectangle indicates the MigrO icons

Phase 1: Trajectory pre-processing. The goal is twofold, to provide insights into the data, and to support data cleaning. Three tools are available: i) the *trajectory plotter* displays the single trajectory as space-time cube enabling the observation of the macro-characteristics of the object's movement; ii) the *summary statistics* tool provides statistics on the trajectories dataset, in particular on the number of points, time and space extent, step length (i.e. distance between two consecutive samples) in space and in time; iii) the *trajectory filter* eliminates the highest sampling frequencies. We recall that the sampling rate has an impact on the clustering operation, therefore controlling the temporal step length is fundamental. Moreover, point filtering can be beneficial for the performance of the clustering algorithm.

Phase 2: **Stay regions generation**. The *presence analysis* tool is a novel functionality that supports the choice of the clustering parameters, in particular of δ (i.e. presence threshold). The tool reports in a graph the number of stay regions as a function of δ (assuming fixed density). The analyst can thus evaluate different scenarios before setting the clustering parameters. The second tool implements the SeqScan clustering algorithm, providing the user with a set of output options.

Phase 3: Stay region analysis. At this stage two tools can be used: the *noise model* allows the extraction of the excursions and transitions performed by the object when residing in a given stay region; the mobility indexes, in particular the *stationarity index graph* and the *commuting index*, for the evaluation and interpretation of clustering.

4. DEMO OUTLINE

We exemplify the analytical process supported by MigrO in two case studies from the domain of animal ecology. The trajectories are from two different deer species that have been equipped with GPS collars and tracked for over one year.

4.1 Seasonal migratory behavior

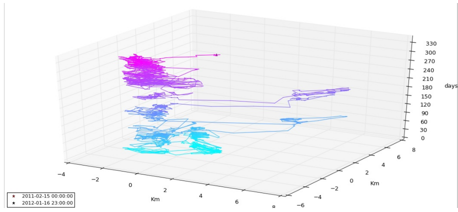
The first set of spatial trajectories are from a group of red deer located in the Bavarian region. This species of ungulates is partially migratory, namely only a portion of the population exhibit migratory behavior. Moreover, migratory behavior can manifest differently across individuals, for example with different distances between seasonal ranges or duration of staying. That motivates the analysis of the animals' trajectories at individual level.

Technically the problem can be re-formulated as follows: for each trajectory, identify the sequence of stay regions where the animal's presence is at least 30 days (for a discussion on the choice of the clustering parameters we refer to [5]). Hence, if the result consists of one stay region, the animal is stationary, otherwise migratory.

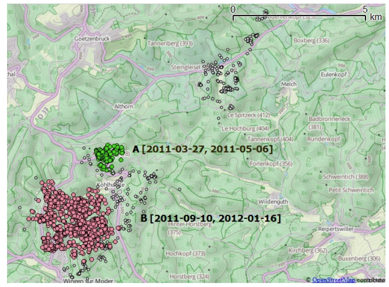
To better illustrate the kind of analysis and the potential of the system, we present a particularly critical case (see Figure 4). From the visual analysis of the trajectory in Figure 4.a one can notice that there are two excursions at two different times and of limited duration while it is unclear whether there is one or two stay regions (typically displayed as large aggregates of points with a significant temporal extent). Figure 4.b shows the outcome of the clustering operation, in particular the two stay regions, labeled A and B. The stationarity index of A is low (< 0.5, omitted), meaning that the animal moves frequently outside the region. The graph of the stationarity index provides details about the movement (when moves in/out) (Figure 4.c). Further, it can be shown, on the basis of the commuting index (omitted for lack of space) that the animal often commutes between A and B, probably due to the presence of an obstacle in between the two regions (i.e. a village).

4.2 Exploratory behavior

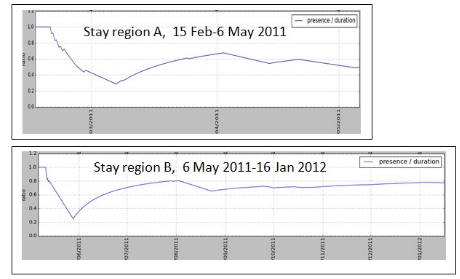
The second set of data refers to a rare experimental occurrence in the wild settings: a group of roe deer were reintroduced in an area, where this species had been extirpated over a century before. Therefore, animals at release were naïve to the study area, and did not have established home ranges. We hypothesize that animal movement after release led to home range emergence, and that this happened through excursions to scope the area, and successive stabilization around preferred sites. MigrO is used as a tool to identify emerging stay regions. To do so, we kept fixed the parameter ϵ (the median step length), adjusted the parameter n to expected density of points from sampling rate and ϵ, and varied the parameter δ, indicating the presence thresh-



(a) Trajectory plotter: the color gradient indicates time evolution



(b) Stay regions: regions A and B as sets of labelled (colored) points



(c) Clustering analysis: stationarity index graph for A and B. The function decreases during excursions and increases or is stable when the object is inside the region

Figure 4: A sketch of the analytical process for the study of the migratory behavior

old. In particular, we varied it from 1.5 days (Figure 5.a) to 10 days (Figure 5.b) and 60 days (Figure 5.c). Doing so, we could observe formation of clusters. Some were abandoned, some instead were increasingly used, until formation by their coalescence of a cluster with longer presence (Fig. 5.c). We consider this as evidence of home-range emergence.

5. CONCLUSIONS

We have presented MigrO, a clustering environment for the extraction of stay regions sequences from spatial trajectories. The platform presents innovative features and covers the key phases of the spatio-temporal clustering. As shown in [5] the SeqScan framework overcomes important limitations of more conventional techniques, offering in particular the opportunity of quantifying the mobility behaviors of individuals.

6. REFERENCES

[1] M.L. Damiani, H. Issa, and F. Cagnacci. Extracting stay regions with uncertain boundaries from from GPS trajectories: a case study in animal ecology. In *Proc. ACM GIS*, 2014.

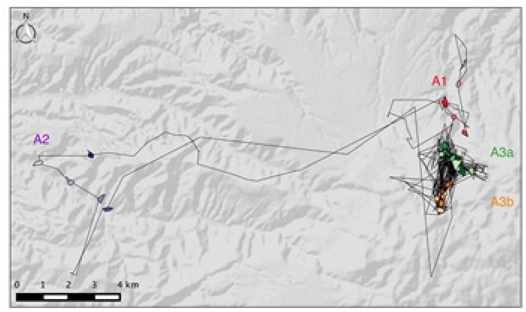
[2] P. Laube. The low hanging fruit is gone: Achievements and challenges of computational movement analysis. *SIGSPATIAL Special*, 7(1), 2015.

[3] S. Dodge, R. Weibel, and A-K Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252, 2008.

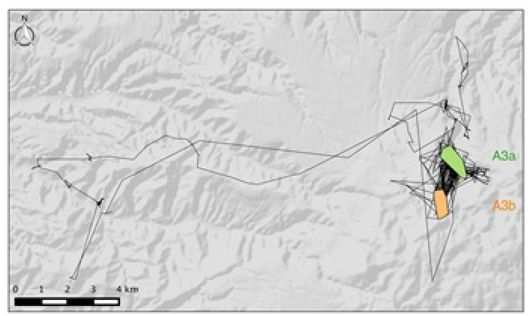
[4] M. Ester, H.P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. KDD*, 1996.

[5] M.L. Damiani, H. Issa, G. Fotino, M. Heurich, and F. Cagnacci. Introducing 'presence' and 'stationarity index' to study partial migration patterns: an application of a spatio-temporal clustering technique. *International Journal of Geographical Information Science*, 29:7, 2015.

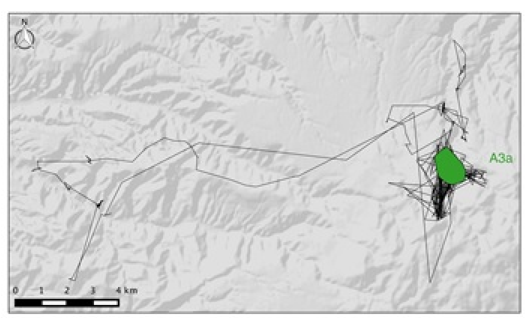
[6] R.H. Güting, F. Valdés, and M. L. Damiani. Symbolic trajectories. *ACM Transactions on Spatial Algorithms and Systems*, Vol. 1, Issue 2:7, 2015.



(a) δ=1.5 days



(b) δ=10 days



(c) δ=60 days

Figure 5: Home-range emergence revealed by coalescence of stay regions at different presence resolution δ. Clusters are noted in four spatially distinct areas A1 (release site), A2, A3a (end of monitoring) and A3b. Home range emergence occurred in area A3a.