Understanding movement in context with heterogeneous data

DOI:
10.1145/3356392.3365222

Document status and date:
Published: 05/11/2019

Document Version:
Accepted manuscript including changes made at the peer-review stage

Understanding Movement in Context with Heterogeneous Data

Onur Derin*
Aniket Mitra
Matei Stroila
onur.derin@here.com
aniket.mitra@here.com
matei.stroila@here.com
HERE Technologies

Bram Custers
Wouter Meulemans
b.a.custers@tue.nl
w.meulemans@tue.nl
TU Eindhoven

Marcel Roeloffzen
Kevin Verbeek
m.j.m.roeloffzen@tue.nl
k.a.b.verbeek@tue.nl
TU Eindhoven

ABSTRACT

Movement data, as captured by myriad sensors, has been growing exponentially. Hence, multidisciplinary approaches for analyzing movement has become feasible. Though, movement pertains to a large variety of domains and applications, the focus of this position paper is understanding human movement (*mobility*) in various forms. We position maps as heterogeneous, multidimensional and digital representation of reality and demonstrate their role in contextualizing movement. We overview the main problems for analyzing human mobility with special attention to movement in context, leveraging heterogeneous data. We review the state-of-the-art in solving these problems and describe remaining open problems and challenges for future work. Finally, we offer a view of existing as well as future mapping and location services that could enable these.

CCS CONCEPTS

• **Information systems → Information systems applications**; **Geographic information systems**.

KEYWORDS

Movement data, context, trajectories, maps

ACM Reference Format:
Onur Derin, Aniket Mitra, Matei Stroila, Bram Custers, Wouter Meulemans, Marcel Roeloffzen, and Kevin Verbeek. 2018. Understanding Movement in Context with Heterogeneous Data. In *MOVE++ 2019: 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data, November 05, 2019 Chicago, IL.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Understanding movement is understanding: (1) its causes (both intrinsic and extrinsic factors), (2) its manifestation (the movement path), and (3) the interplay between the two, causes and manifestation. A good introduction to understanding movement is by Dodge et al. [8] in which they define movement as spatio-temporal signals, which carry information about the movement of dynamic entities and the underlying mechanisms that drive their movement. Moreover, movement, and more specifically geospatial movement, can be perceived from two different perspectives [13]: (1) the Lagrangian view, which considers changes in a moving objects location, and (2) the Eulerian view, which describes changes in a moving objects location with respect to fixed points in space.

Various techniques are used to track location of people, animals, vehicles and other objects. This ubiquitous application of location sensors has led to an ever-increasing volume of data regarding the geospatial location of such entities. GPS is a prime example of this and is extensively used to track human movement (*mobility*) in various forms (walking, driving, bicycling, etc.). Most computational techniques for understanding mobility focus almost exclusively on such trajectory data (location over time per entity). Yet, mobility is influenced by various factors or *context*, both internal (e.g. motivation, capabilities) and external (e.g. terrain, other entities) to the moving entity, which enable and limit movement. Thus, taking this context into account allows for improving the accuracy of mobility analysis as well as opening up new analysis possibilities [10]. Here, we focus on external context.

The inclusion of context makes data highly heterogeneous. Buchin et al. [2] identify context as the location circumstances of a moving agent, which includes external factors connected to the underlying landscape (e.g., the road network) or the surrounding environment (e.g., gas-stations, drive-throughs). Additionally, temporal context (e.g., time of day or when a point of interest is open) can also influence movement.

Heterogeneous data provides a multidimensional view of both movement and context [8]. Hence, it is important to distinguish between heterogeneous movement data and heterogeneous context data. The former requires collection, storage and fusion of movement data collected from the Lagrangian and the Eulerian perspective. This would mean, for example, collection and fusion of a GPS trajectory of an entity together with its movement recorded by beacons installed at multiple locations (e.g. [17, 19]). The latter on the other hand can be combined to model the context that drives movement. For example, social signals that describe the popularity of a point of interest (POI) combined with availability of parking facilities around the POI can be fused to form a context that enables us to understand and analyze movement around the POI.

Abundance of data also enables for adoption of data-driven approaches in understanding and predicting movement in context [7]. A pragmatic approach would be to (1) use computational

*All authors contributed equally to this research.

MOVE++ 2019, November 05, 2019, Chicago, IL

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

movement-analysis techniques as a preprocessing step to data-driven approaches or (2) use data-driven models for better parameter estimation of model-based computational movement-analysis techniques.

Maps and location-based services have always been a means for finding locations of interest and navigating to those locations. Traditionally, maps have added context to movement by defining the underlying landscape that enables or limits movement. However, with the availability of heterogeneous context data, maps can take on a much bigger role in contextualizing movement.

Our aim with this position paper is three-fold: 1) identify classes of problems for human movement in context, 2) offer a glimpse at the state-of-the-art work in solving these and 3) describe some of the remaining open problems and challenges for future work. We position the role of *maps*[1] as a digital abstraction of reality in understanding mobility, by contextualizing movement. Finally, we offer a view of existing as well as future mapping and location services that could enable these.

2 MAPS AS CONTEXT

As a computational structure for navigation, maps traditionally consist of a labeled geometric graph modeling the road network. Movement is then contextualized by map-matching, wherein each point of the movement path is matched to an edge of the graph. The result of the map-matching process is contextually enriched movement data with additional attributes such as distance from the nearest edge and the attributes of that edge.

We extend the idea of *map* to be a general digital representation of context. Similarly, we then conceptually extend map-matching to use a broader class of context for contextualizing movement. As shown in Figure 1, such a map may have many facets, for example: human artifacts (roads, buildings, etc.), natural features (terrain elevation, land cover, etc.) and ambient measures (temperature, wind velocity, etc.). Map features can be categorized according to various aspects. Some aspects are inherently discrete (e.g., roads, buildings), whereas others model continuous phenomena (e.g., temperature). Additionally, features can be static or dynamic, that is, they may change over time. Change may happen in various forms (features may appear, disappear, move or take on different attribute values) and happen at different rates (terrain elevation rarely changes, the road network changes more often, whereas temperature changes constantly). Features influence movement in different ways. Some may naturally enable movement (road network), whereas others limit movement (heavy precipitation). This even depends on the mode of transport: water features limit car movement, but enable movement by boats.

In this regard, we select some algorithms used in computational movement analysis that can potentially benefit from context as input, highlight some state-of-the-art work and also present open problems and possible directions for future work.

2.1 Data Correction: Outlier Detection

Movement data is often prone to errors, for example GPS trajectories show a number of erroneous recordings mainly in urban canyons, thereby resulting in outliers. Therefore, identification and

[1] In our terminology, a map is a data model, not a cartographic map.

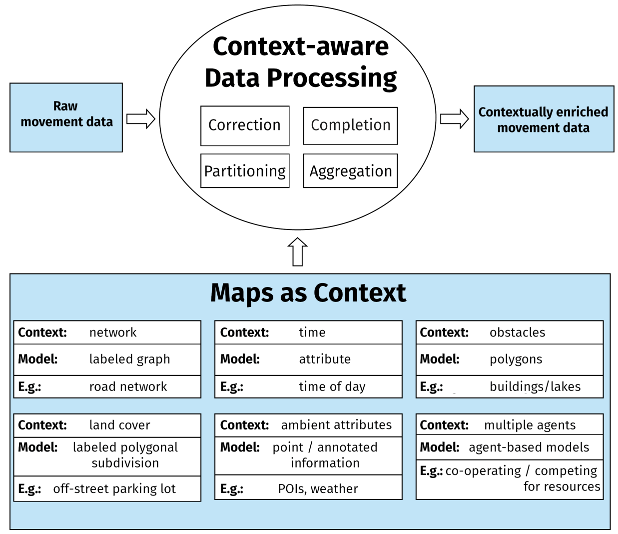

Figure 1: Contextualizing Movement

removal of outliers is a fundamental preprocessing step in computational movement analysis. A number of techniques for outlier detection and removal have been proposed in literature which take into account the kinematics of the movement to identify outliers [5]. A context-aware outlier detection algorithm would take in raw movement data and generate contextually sanitized data.

Open problem 1. *How can we use context, such as speed limits and traffic flow, to improve outlier-detection algorithms based on movement models?*

Open problem 2. *Can we use machine learning techniques for detecting outliers by integrating context? Which type of context is relevant for such methods?*

2.2 Data Completion: Interpolation

Long [14] incorporates object kinematics (i.e., velocity and acceleration) into the interpolation process. However, people do not navigate alone in a given space. Their navigation capacity (e.g., the speed at which they move) is determined by the space itself and how other people move in that space at the same time. Hence, context (e.g., the speed limit and the traffic flow along a road segment should play an important role in interpolation.

When context is applied to interpolation, a natural extension is spatio-temporal interpolation. Spatio-temporal interpolation has its roots in geo-statistics. However, one of its limitations is that it requires movement to be modeled as a stochastic process. While such a modeling approach may result in a good representation of a trajectory as a whole, context has a very localized influence on a trajectory. Therefore there is a need for better models for spatio-temporal interpolation that takes context into account. A context-aware interpolation algorithm would take in movement data with gaps and fill in the gaps such that it results in a trajectory which is more representative of the context in which the movement took place.

Open problem 3. *How can context-aware spatio-temporal interpolation algorithms be designed that take in movement data with gaps and fill in the gaps?*

2.3 Data Partitioning: Segmentation

Buchin et al. [3] describe segmentation as partitioning a trajectory into sub-trajectories such that the movement characteristics within each sub-trajectory is uniform in some sense. They propose a criteria-based segmentation wherein the criteria can be kinematic (e.g., speed, acceleration and heading) or contextual (e.g., shape fitting). If the shape is a collection of links that model the underlying road network, then the authors propose a criteria that can be used to partition a trajectory if it is similar to the given shape.

Alewijnse et al. [1] model movement as a dynamic Brownian bridge movement model (dBBMM) and propose a model-based segmentation algorithm which minimizes the information criteria among all possible segmentations. Defining movement as a model captures the context of movement. However, dBBMM has been used to model animal movement and is not necessarily a good choice to model human mobility.

The approach by Sila-Nowicka et al. [18] integrates GPS trajectories with contextual information. They partition the trajectory into homogeneous segments that correspond to different movement modes. Furthermore they use machine learning techniques to classify each segment as *stop*/*movement* and further classify each *movement* segment as *walk*/*vehicle*. Finally, they propose to associate an activity label to each "stop" segment by using contextual information like proximity of the segment to a POI.

A context-aware segmentation algorithm would take in movement data and generate contextual partitions of it.

Open problem 4. *Which movement model best describes human mobility? Can such movement models benefit from supervised learning for estimation of their parameters?*

2.4 Data Aggregation: Clustering

An important tool in dealing with large volumes of data is the aggregate. In the context of movement analysis, this often comes in the form of trajectory clustering. This allows us to analyze not just one moving entity but a large collection, but also to understand a single entity in the context of multiple moving entities. This is relevant both for movement on the road network, but also other forms of movement (e.g., pedestrian movement, or off-network movement near POIs and off-street parking).

Buchin et al. [2] propose methods to compute similarity of trajectories by taking into account geographical context. In this work they consider context as a labeled polygonal subdivision and define context distance as the geodesic shortest path between two vertices of the dual graph of the planar subdivision. Furthermore, they show how to combine the context distance with a similarity metric for trajectories (e.g., Fréchet distance). Kostitsyna et al. [11] describe a method of detecting groups of moving entities, by identifying entities that are spatially close during some time interval. Their method takes context into account in the form the geodesic distance that avoids polygonal obstacles.

Open problem 5. *How can we differentiate or cluster trajectories in a general heterogeneous context? Which forms of context are most relevant for such clustering task, and how does this relate to the type of moving entities?*

2.5 Routing

The above problems relate to data processing and enriching movement data based on context. Various other computational problems can use such enriched data or use context directly. Here, we illustrate this using two routing problems.

Building accurate models for understanding the flows in a spatio-temporal network is a very important problem for predicting people's mobility and designing complex multi-modal transportation solutions to reduce traffic congestion and incidents. The flows are affected by multiple complex factors, such as spatial correlations between different locations, temporal correlations among different time intervals, and external factors such as social events, weather, road closures and incidents. A context-aware routing can benefit from a traffic prediction model that takes into account external factors. Zhang et al. [20] takes a step towards such a solution by proposing a multi-task deep-learning framework that integrates external factors and simultaneously predicts the flows, but their prediction is limited to flows in a grid representation of a city.

Open problem 6. *Can we predict traffic at the road segment level by taking context into account?*

As described in section 2, a map that contextualizes movement, also treats multiple agents as context. In agent-based models, the movements of these agents is driven by the context defined by (a) the network, time, ambient attributes like weather, obstacles such as incidents etc. and (b) other agents. The context of other agents can facilitate co-operative or competitive movement. An example for co-operative movement can be vehicles on the highway allowing other vehicles to merge. Another well-studied problem is that of managing congestion in networks, as surveyed in [6], for which knowledge of the agents in the system is crucial. Similarly, a competitive movement can emerge when a large number of vehicles try to get to a parking place that is nearest to their destination. This is predominantly a problem in downtown areas and can lead to extra costs in time and fuel. Moreover, there may be scenarios in which movement can be both co-operative and competitive. For example, in case of routing during an emergency, movement can switch from co-operation to competition in the absence of the right control policy. The entity that controls the agents or the resources has a global objective function that it wants to optimize while each agent might have its own objective function which may not be necessarily aligned with the objective function of the controlling entity. Therefore there is a duality between how a controlling entity expects movement to manifest vs. how the agent actually moves.

Open problem 7. *How do multiple agents contextualize co-operative vs. competitive movement and what control policies facilitate co-operative movement?*

3 PRIVACY AND CONTEXT ENRICHMENT

Understanding movement requires data in the form of GPS trajectories combined with additional information such as biometric data (if people), CAN bus and video data (if vehicles). Some companies or organizations may possess such data, allowing (computational) analysis. However, availability of data at scale for the entire scientific community (hence, open data) is sparse due to privacy concerns. Attempts to open datasets often result in potentially disclosing personal information [4].

Krumm [12] showed that it is possible to infer the coordinates of a subject's home and identity from pseudonymous GPS data. In order to carry out such an inference attack, the author segmented the trajectories and attached context to these segments. For instance, four heuristic techniques were used to identify a subject's home, namely (1) identifying destination clusters formed by end segments of the trajectory within a certain temporal range (2) identifying destination clusters formed by longest dwell time (3) merging nearby destination clusters as a super cluster (4) identifying a spatio-temporal distribution that gives the probability of a subject being in a given destination at different times of the day (e.g., day vs. night) with possible destinations obtained by reverse geo-coding the GPS data.

One solution to prevent inference attacks as demonstrated in [12] is to generate synthetic movement data that is statistically indistinguishable from real data. Ouyang et al. [15] present a non-parametric generative model of human mobility. This model synthesizes realistic human trajectories that preserve not only the statistical characteristics of the trajectories used for training, but also their intrinsic semantics. They use a non-sequential approach, where the entire trajectory at once is taken in consideration as input. Thus, the model not only finds the correlation between visited locations but also learns the common semantic/geographic patterns of mobilities across the training data.

Open problem 8. *Contextually enriched movement data also increases the risk of personal information being inferred. Data anonymization algorithms need to be designed and analyzed that can prevent such attacks. Which inference attacks can be used to expose vulnerabilities in such data? Which data anonymization algorithms prevents inference attacks on contextually enriched movement data?*

4 PLATFORM FOR LOCATION DATA

To facilitate analysis and prediction of context-aware movement, there is a need for a platform that contains rich context data. Such a platform can (a) ingest and model heterogeneous context data, (b) facilitate combining and manipulating different context models, and (c) allow the context models to be used with different types of movement analysis and prediction models.

While traditional maps are road-centric both in content and use cases, the newer/future mapping services offer a location platform that can cope with massive amounts of contextual, heterogeneous geodata, and can use it to produce enhanced models and predictions for location intelligence, mobility and autonomy applications.

As movement analysis is an interdisciplinary research field, a standard movement-analysis/-prediction toolset could serve to combine various techniques in a coherent manner. While some libraries such as *MovingPandas* [9] and *scikit-mobility* [16] exist, they do not encompass the entire spectrum of analysis and prediction techniques. For instance, algorithms based in computational geometry have not been covered by these tools.

Open problem 9. *How can we build a reusable, high performance, feature-rich library that incorporates state-of-the-art computational geometry algorithms for computational movement analysis?*

5 CONCLUSIONS

In this position paper, we identified several classes of problems for human movement in context, offered a glimpse at the state-of-the-art work in solving these and described some of the remaining open problems and challenges for future work. Along these lines, we positioned the role of the maps, via their extended definition as a digital representation of context, in understanding mobility by contextualizing movement.

ACKNOWLEDGMENTS

This work was supported by HERE Technologies and the Netherlands Organisation for Scientific Research (NWO) under project no. 628.011.005.

REFERENCES

[1] S. P. A. Alewijnse, K. Buchin, M. Buchin, S. Sijben, and M. A. Westenberg. 2018. Model-Based Segmentation and Classification of Trajectories. *Algorithmica* 80, 8 (2018), 2422–2452.
[2] M. Buchin, S. Dodge, and B. Speckmann. 2014. Similarity of trajectories taking into account geographic context. *J. of Spatial Information Science* 9 (2014), 101–124.
[3] M. Buchin, A. Driemel, M. J. van Kreveld, and V. Sacristán Adinolfi. 2011. Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. *J. of Spatial Information Science* 3 (2011), 33–63.
[4] C. Culnane, B. I. P. Rubinstein, and V. Teague. 2019. Stop the Open Data Bus, We Want to Get Off. arXiv:cs.CR/1908.05004
[5] B. Custers, M. van de Kerkhof, W. Meulemans, B. Speckmann, and F. Staals. 2019. Maximum Physically Consistent Trajectories. In *Proc. 2019 ACM SIGSPATIAL GIS*. to appear.
[6] P. Desai, S. W Loke, A. Desai, and J. Singh. 2011. Multi-agent based vehicular congestion management. In *IEEE Intelligent Vehicles Symp. (IV)*. IEEE, 1031–1036.
[7] S. Dodge. 2019. A Data Science Framework for Movement. *Geographical Analysis* to appear (2019), 1–21.
[8] S. Dodge, R. Weibel, S. C. Ahearn, M. Buchin, and J. A. Miller. 2016. Analysis of movement data. *Int. J. of GIS* 30, 5 (2016), 825–834.
[9] A. Graser. 2019. MovingPandas: Efficient Structures for Movement Data in Python. *GI_Forum* 1 (2019), 54–68.
[10] C. Gschwend and P. Laube. 2012. Challenges of Context-Aware Movement Analysis - Lessons Learned about Crucial Data Requirements and Pre-Processing. In *Proc. 20th GIS Research UK Annual Conference*. 241–246.
[11] I. Kostitsyna, M. van Kreveld, M. Löffler, B. Speckmann, and F. Staals. 2015. Trajectory Grouping Structure under Geodesic Distance. In *Proc. 31st Int. Symp. on Comput. Geom.* 674–688.
[12] J. Krumm. 2007. Inference Attacks on Location Tracks. In *Proc. Int. Conf. on Pervasive Computing (LNCS 4480)*. 127–143.
[13] P. Laube. 2014. *Computational Movement Analysis*. Springer Publishing Co., Inc.
[14] J. A. Long. 2016. Kinematic interpolation of movement data. *Int. J. of GIS* 30, 5 (2016), 854–868.
[15] K. Ouyang, R. Shokri, D. S. Rosenblum, and W. Yang. 2018. A Non-Parametric Generative Model for Human Trajectories. In *Proc. 27th Int. Joint Conf. on Artificial Intelligence*. 3812–3817.
[16] L. Pappalardo, G. Barlacchi, F. Simini, and R. Pellungrini. 2019. scikit-mobility: An open-source Python library for human mobility analysis and simulation. *arXiv e-prints* (Jul 2019), arXiv:1907.07062.
[17] M. Rahmani, E. Jenelius, and H. N Koutsopoulos. 2014. Floating car and camera data fusion for non-parametric route travel time estimation. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 1286–1291.
[18] K. Siła-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. Demšar, and A. S. Fotheringham. 2016. Analysis of human mobility patterns from GPS trajectories and contextual information. *Int. J. of GIS* 30, 5 (2016), 881–906.
[19] M. Wright and R. Horowitz. 2016. Fusing loop and GPS probe measurements to estimate freeway density. *IEEE Trans. on Intelligent Transportation Systems* 17, 12 (2016), 3577–3590.
[20] J. Zhang, Y. Zheng, J. Sun, and D. Qi. 2019. Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning. *IEEE Trans. on Knowledge and Data Engineering* to appear (2019), 1–11.