Latent Terrain Representations for Trajectory Prediction

Andrew Feng
Andrew S. Gordon
feng@ict.usc.edu
gordon@ict.usc.edu
Institute for Creative Technologies, University of Southern California
Los Angeles, California

ABSTRACT

In natural outdoor environments, the shape of the surface terrain is an important factor in selecting a traversal path, both when operating off-road vehicles and maneuvering on foot. With the increased availability of digital elevation models for outdoor terrain, new opportunities exist to exploit this contextual information to improve automated path prediction. In this paper, we investigate predictive neural network models for outdoor trajectories that traverse terrain with known surface topography. We describe a method of encoding digital surface models as vectors in latent space using Wasserstein Autoencoders, and their use in convolutional neural networks that predict future trajectory positions from past trajectory data. We observe gains in predictive performance across three experiments, using both synthetic and recorded trajectories on real-world terrain.

CCS CONCEPTS

• **Information systems → Geographic information systems**; • **Computing methodologies → Dimensionality reduction and manifold learning**.

KEYWORDS

neural networks, trajectory prediction, latent representation

ACM Reference Format:
Andrew Feng and Andrew S. Gordon. 2019. Latent Terrain Representations for Trajectory Prediction. In *1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE'19), November 5, 2019, Chicago, IL, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3356392.3365218

1 INTRODUCTION

In the task of predicting a person's future trajectory given recent observations, the surrounding environmental context can be informative. When walking in urban settings, for example, people may be more likely to follow the route of sidewalks, or keep a consistent distance from the walls of buildings. Likewise, when traversing wilderness areas, people will adjust their paths in consideration of the features of the terrain, based on how quickly, safely, or efficiently they aim to reach their intended destination. With the increased availability of digital surface models for arbitrary locations on Earth, new opportunities exist for using the physical shape of the terrain improve prediction for GPS-encoded paths. Needed are methods for representing terrain features for arbitrary locations, and an approach to path prediction that can directly exploit the information in these representations for improved accuracy.

As in other prediction tasks, machine learning methods based on neural networks are well-suited to path prediction, and offer several approaches to the integration of terrain information. In the simplest approach, network architectures could include the nearby raw digital surface model as an additional set of input features alongside the available time-series GPS data. With infinite training data and computational resources, such an architecture could learn generalized features of the terrain that are most informative. In seeking a more practical approach, we note the relative discrepancy in the size of available geospatial terrain data (huge) compared to GPS-encoded path data (modest). Given the imbalance, we see an opportunity to divide the machine learning task into two parts, first learning task-independent terrain representations from the ubiquitous data, then exploiting them in the specific task of path prediction.

In this paper, we investigate an approach to neural path prediction incorporating latent terrain representations, divided into two parts. First, we train a low-dimensional encoder for raw height values in digital surface model data using a Wasserstein Autoencoder. Second, we train a trajectory prediction model using convolutional neural networks, concatenating input trajectory offsets with encoded representations of their surrounding terrain. We evaluate the prediction accuracy of our model as compared with an ablated model trained using only the trajectory offsets. Three datasets are used for our evaluation, consisting of synthetic data, recorded data of hikers, and recorded data from both vehicles and people on foot in a desert military exercise.

2 RELATED RESEARCH

The analysis of path information in wilderness areas is an active area of ecology research, where GPS tracking devices have been attached to animals in a wide variety of studies for different purposes. Traditionally seen as a tool for habitat range identification, paths from high-frequency sampling of animal location create new opportunities to study other aspects of behavior such as habitat use and diet [Weber et al. 2015]. In analyzing animal path data, a central concern in ecology research is path segmentation rather than prediction [Edelhoff et al. 2016], typically for the purpose of

MOVE'19, November 5, 2019, Chicago, IL, USA

ACM ISBN 978-1-4503-6951-0/19/11…$15.00
https://doi.org/10.1145/3356392.3365218

identifying the behavioral state of the tracked animal. State-space models for the analysis of animal tracking data [Patterson et al. 2008] have some similarities to the neural path prediction approach we pursue in this paper, as their latent encoding of intrinsic context (behavioral state) is analogous to our latent encoding of extrinsic context (terrain).

Learning a useful representation and concentrating the data on low dimensional manifolds are important for machine learning tasks. For self-supervised learning, the classical auto-encoder produces a lower-dimensional representation by minimizing reconstruction errors of original data. The pioneering work of variational autoencoder [Kingma et al. 2014] aims to improve the latent space by adding a regularizer that minimizes the KL-divergence. It produces a more structured latent space with better interpolation results in the unseen areas. Wasserstein Autoencoder [Tolstikhin et al. 2018] proposes a regularizer with optimal transport divergence between two distributions. It matches the data distribution in latent space with a prior distribution, and can generate samples with better reconstruction quality.

Trajectory prediction aims to predict future movements based on past trajectories. However, using only the previously observed movement as input ignores the factors contributed by surrounding environments. The work in SocialLSTM [Alahi et al. 2016] assumes that the nearby pedestrian movements also affect the future trajectory of an agent. The follow-up work incorporates nearby scene information implicitly by using common historic movements within a grid cell to predict future trajectory [Manh and Alaghband 2018]. The work by Zhang et al [Zhang et al. 2018] integrates low-level LiDAR statistics to learn feature representations for motion prediction.

3 METHODOLOGY

The main issue in trajectory prediction based only on the observed path in the past is that the given information may not be enough to accurately model the future path. For example, if the observed trajectory is heading toward a fixed direction, there is little information to indicate whether the future path will continue in the same direction or making a turn at some point. We propose to improve this information gap by adding nearby terrain information into the prediction of future trajectory.

Our method is separated into two steps: terrain latent space generation and trajectory prediction. For the first stage, we apply Wasserstein Autoencoder [Tolstikhin et al. 2018] to obtain the latent space from nearby terrain digital elevation models (DEM). The latent variables obtained from the DEM patch centered at each trajectory point are then concatenated with the corresponding trajectory coordinates to form the input vector for the next stage. For the second stage, instead of using RNNs such as LSTM, we learn a convolutional neural network to predict the future trajectory positions from past trajectory and corresponding terrain latent space.

3.1 Terrain Latent Space Generation

Based on all of the trajectory points in the training dataset, we calculate the bounding area in latitude/longitude and then obtain its corresponding DEM. The DEM imagery is obtained from Bing map [Rischpater and Au 2013] and we use the zoom level 13, which equates to 19 meters per pixel. Then a sliding window of 64 by 64 pixels is used to sample DEM patches from the terrain DEM, which covers about the 1.2km by 1.2km area. These patches are then used to train a Wasserstein Autoencoder (WAE) to compute a latent space representation for the terrain. Here we set the size of the latent vector to 8. Compared to Variational Autoencoder (VAE), WAE tend to produce a sharper image reconstruction, and thus can preserve the local geometry changes better.

To help visualize what is being learned by the WAE network, we trained the model on DEM data from the deserts of Southern California, and applied a four-color overlay representing clusters of similarly-represented patches, as shown in Figure 1. To assign each patch to one of the four clusters, the k-means clustering (where K=4) is applied on DEM patches using the Euclidean distances of the height values in each pixel or latent vectors, shown on the left and the right sides of Figure 1, respectively. We observe that clustering with raw values distribute patches based on absolute height values and thus produce large clusters across both sloped and flat areas. On the other hand, latent vector encodes local geometry features better and the resulting clusters tend to align with local slope directions.

3.2 Trajectory Prediction

Using the learned WAE, a corresponding DEM patch centered at each trajectory point is extracted and its latent vector is computed. The latent vectors are then concatenated with trajectory offsets to form the input data. To train the trajectory prediction model, we use the CNN architecture similar to the one proposed in [Nikhil and Morris 2019] instead of a typical recurrent neural network such as LSTM. Recent research has shown that convolutional architecture could outperform the recurrent neural network for modeling sequence data [Bai et al. 2018] and therefore we choose convolutional network as our starting point for experiments. Specifically, the network contains 4 layers of 1D CNN with kernel size 10 and output channel size 32, followed by a fully connected layer with output size 20. The network is trained with the mean absolute error loss using the RMSProp optimizer. To produce training data, the offset vector from the previous time step is first computed for each trajectory at each time step. Then we use a sliding window of size 30 across these offsets to extract samples with equal size and apply zero padding for the missing data. For each sample, the offsets and terrain latent vectors from the first 10 time steps will be used as the input and the trajectory offsets for the next 20 time steps are the output. At testing time, the predicted offsets are then used to reconstruct trajectory positions.

4 EXPERIMENTS

We use three different data set to test our trajectory prediction model with terrain latent space. Among them, one is a synthesized trajectory dataset and the other two are real-world datasets. For each dataset, we train on 75% of the trajectories and validate on the remaining 25%.

4.1 Synthetic Trajectories

The synthetic data is created by following the slope descending direction from randomly selected starting points in the terrain. To

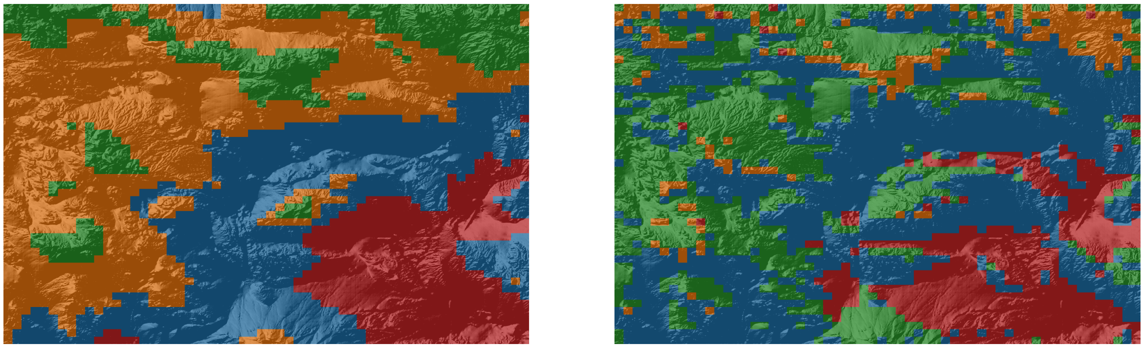

Figure 1: K-means clustering of terrain patches based on height values at each pixel (left) and latent representations (right). The terrain covers the area of approximately $60 \times 55\ km^2$.

Figure 2: Examples of trajectories used in the experiments from the Synthetic dataset (left), the Hiking dataset (middle), and the Desert dataset (right). The synthetic trajectories follow the descending directions and are highly correlated to the nearby terrains. The Hiking and Desert trajectories consist of recorded GPS tracks of hikers and military units (infantry and vehicles), respectively.

generate the data, we choose the terrain DEM data from the region of Ibiza island, Spain (N39°1′12.04″ E1°28′55.73″). The goal is to create trajectories that will have a strong correlation to nearby terrain environments to validate the usage of terrain latent space for prediction. The examples of synthesized trajectories are shown in Figure 2. As shown in Table 1, using latent space terrain as part of input in the model improves the prediction accuracy significantly in the synthetic dataset. This validates the usage of terrain information when the trajectories are correlated with nearby geometries.

4.2 Real-World Trajectories

We selected two GPS trajectory datasets collected from rugged (non-plain) areas to test the terrain latent space for trajectory prediction, illustrated in Figure 2. The Hiking trajectory data is obtained from [Lera 2017], which is a collection of 15,376 GPS tracks of hikers in the Balearic Islands of Spain from 2009 to 2016, collected from approximately 2,000 participants. The sample rate for this dataset was about 0.04Hz to 0.06Hz. We use a subset of the data containing 1554 trajectories in our experiment.

The Desert trajectory data is collected during a military training exercise held at the National Training Center at Ft. Irwin, California using the Multiple Integrated Laser Engagement System (MILES) [Data 2019]. These GPS trajectories were obtained from approximately 500 infantry and vehicle units, collected during 90 minutes of the training exercise. The sample rate for this dataset was between 0.03Hz to 0.08Hz. To clean up these non-uniform sampling rates, we resampled both datasets to 0.1 Hz as a pre-processing step.

For each GPS coordinate in each track of each dataset, we computed the latent representation of the surrounding terrain by cropping a 64 by 64 pixels DEM patch centered at that location. The WAE for each dataset is trained separately using the DEM data from that region for simplicity. With enough DEM training images, a single WAE could be trained to represent terrain feature space across all datasets. The DEM data is obtained via Microsoft Bing Map [Rischpater and Au 2013] with detailed level 13.

For each track in each test split, we predict the locations of the next 20 timestamps that follow an input sequence of 10 timestamps, and compute two accuracy metrics. The first is average offset error, which is the average error between the predicted offset vector and the ground truth offset at each time step, and the other is average displacement error, which is the mean-square error between predicted points to the true locations, following the previous experimental paradigm of Social LSTM [Alahi et al. 2016].

Table 1 shows the comparison prediction accuracy with and without latent space terrain. Accuracy gains in all three evaluations are significant ($p < 0.001$), using stratified shuffling [Yeh 2000]. Although the improvements seen in the two real-world datasets are not as substantial as in the Synthetic dataset, the results consistently show around 5% decrease in mean-square errors when using terrain information. Figure 3 shows a visual comparison example between the two methods for the Hiking data. Using latent space terrain benefits the path prediction the most near curvy turns, where there are not enough information in the observed trajectory to predict such a sharp turn.

5 DISCUSSION

The improvements observed in the aforementioned experiments indicate that latent representations of terrain can improve the prediction of future trajectories of people in outdoor, wilderness settings. Unsurprisingly, improvements are most substantial in our experiments with synthetic data, since nearby terrain is highly correlated with the synthesized movements. Primarily, this experiment with synthetic data serves to validate that the latent space

Table 1: Accuracy of trajectory prediction, reported as mean-square error (lower is better). Both the reported offset and displacement errors are in meters. All improvements are significant (p < 0.001).

Dataset	Size	Trajectory Only		Trajectory + DEM		Improvement	
		Offset	Displacement	Offset	Displacement	Offset	Displacement
Synthetic	500	0.3647	23.848	**0.2763**	**13.405**	24.23%	43.78%
Hiking	1554	0.2887	19.926	**0.2757**	**18.992**	4.48%	4.68%
Desert	1332	1.0035	84.556	**0.9535**	**79.119**	4.99%	6.45%

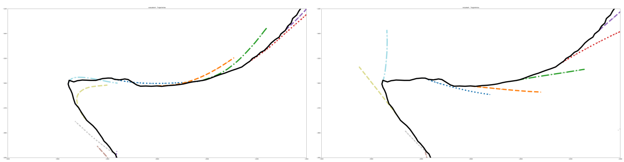

Figure 3: Comparison of trajectory prediction with (Left) and without (Right) latent space. Overall using latent space helps to predict some sharp turns where the information from partly observed trajectory is not enough to predict the turn.

learned by the WAE is, indeed, encoding terrain features that are descriptive of its surface shape. Our two experiments with recorded, real-world trajectories also show reduced prediction error when incorporating latent representations of terrain, albeit with smaller improvements. We find it somewhat surprising that the improvements in the Desert dataset were more pronounced than in the Hiking dataset, as it included more variation in the types of tracked entities (both mounted and dismounted military forces). We note, however, that the mean square error is substantially lower in the Hiking dataset in both conditions. This suggests more uniformity in the types of paths traversed by these hikers across the dataset, as well as the pace and behavior of the hikers themselves.

While our experiments validate the inclusion of latent terrain representations in trajectory prediction, further investigations are needed to better understand how these representations can best be exploited. In this work, we choose to train the WAE as a separate process, before training the CNN network. Combining the terrain encoding and trajectory prediction into the same network would allow end-to-end training, and could potentially improve the accuracy of the final network. We also believe that alternative network architectures, e.g., LSTM, transformers, should be investigated. As well, further insights can be gained through experimentation with a wider range of trajectory datasets and varied terrain. In addition to human movement datasets, tracks of wild animal movements are also likely to be highly correlated to nearby terrain features, opening up new applications of our approach in ecology research and wildlife management. Although the majority of public datasets we could find from the open repository [Bank 2017] do not record GPS coordinates at sufficiently high sampling rates for use in trajectory prediction, new datasets with high sampling rates will open up new opportunities for future research.

ACKNOWLEDGMENTS

The projects or efforts depicted were or are sponsored by the U. S. Army. The content or information presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social LSTM: Human Trajectory Prediction in Crowded Spaces. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).

Move Bank. 2017. A Free, Online Database of Animal Tracking Data. https://www.movebank.org/.

MILES Data. 2019. Multiple integrated laser engagement system. https://en.wikipedia.org/wiki/Multiple_integrated_laser_engagement_system.

Hendrik Edelhoff, Johannes Signer, and Niko Balkenhol. 2016. Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. *Movement Ecologyvolume* 4 (2016).

Diederik P. Kingma, Danilo J. Rezende, Shakir Mohamed, and Max Welling. 2014. Semi-supervised Learning with Deep Generative Models. In *Advances in Neural Information Processing Systems 27 (NIPS), Montreal, 2014.*

Isaac Lera. 2017. Dataset of Human Hiking GPS Trajectories on Balearic Islands (SPAIN). https://osf.io/77n9t/.

Huynh Manh and Gita Alaghband. 2018. Scene-lstm: A model for human trajectory prediction. *arXiv preprint arXiv:1808.04018* (2018).

Nishant Nikhil and Brendan Tran Morris. 2019. Convolutional Neural Network for Trajectory Prediction. In *Computer Vision – ECCV 2018 Workshops*, Laura Leal-Taixé and Stefan Roth (Eds.). Springer International Publishing, Cham, 186–196.

Toby A. Patterson, Len Thomas, Chris Wilcox, Otso Ovaskainen, and Jason Matthiopoulos. 2008. State-space models of individual animal movement. *Trends in Ecology and Evolution* 23 (2008), 87–94. Issue 2.

Ray Rischpater and Carmen Au. 2013. *Microsoft Mapping: Geospatial Development with Bing Maps and C#* (1st ed.). Apress, Berkely, CA, USA.

Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. 2018. Wasserstein Auto-Encoders. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings.*

Natalie Weber, Prateep Duengkae, Jakob Fahr, Dina K. N. Dechmann, Patcharakiti Phengsakul, Wachirapon Khumbucha, Boripat Siriaroonrat, Supaporn Wacharapluesadee, Pattarapol Maneeorn, Martin Wikelski, and Scott Newman. 2015. High-resolution GPS tracking of Lyle's flying fox between temples and orchards in central Thailand. *The Journal of Wildlife Management* 79, 6 (2015), 957–968.

Alex Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th Conference on Computational Linguistics.* 947–953.

Yanfu Zhang, Wenshan Wang, Rogerio Bonatti, Daniel Maturana, and Sebastian Scherer. 2018. Integrating kinematics and environment context into deep inverse reinforcement learning for predicting off-road vehicle trajectories. In *Conference on Robot Learning.* 894–905.