Spatiotemporal Learning of Directional Uncertainty in Urban Environments

Weiming Zhi[1] Ransalu Senanayake[2] Lionel Ott[1] Fabio Ramos[1,3]

***Abstract*— We present a model that captures the long-term motion directions of dynamic objects, such as pedestrians. We approach this problem by modelling movement directions of a typical object in an environment over time. At a given coordinate in space and time, our method provides a multi-modal probability density function over the possible directions an object can move in. The model is continuous in the spatial and temporal domains, as we can query the change of directional uncertainty at arbitrary resolution in space and time.**

I. Introduction

Autonomous robots may need to operate in urban environments with moving objects, such as crowds of people. The understanding of movement directions can shed light on representing the trajectories of these dynamic objects in the environment. This work addresses the problem of understanding trajectories of dynamic objects by building a directional probabilistic model that is continuous over time and space.

The method provided in our work is aimed at learning the long-term dynamics in an environment. There have been attempts [1], [2] to extend occupancy mapping beyond static environments by storing occupancy signals over time, and building a representation along time in each grid. Instead of modelling a changing occupancy map over time, methods [3] have also been developed to understand long-term occupancy by that capture movement directions in the environment.

Our method builds a model with the following desirable properties:

1) models the probability distribution of movement directions over a valid support of $[-\pi, \pi)$, and the distribution can be multi-modal;
2) represents the environment in a continuous manner, without assuming the discretisation of the environment into a grid of fixed resolution, with independent cells;
3) takes into account how the probability distribution of directions changes over time.

II. Method

A. Problem Formulation

We consider a dataset of $\mathcal{D} = \{(x_m, y_m, t_m, \theta_m)\}_{m=1}^{M}$. x_m and y_m denote the longitude and latitude of the data point in space, t_m denotes the time step at which the m^{th} observation was made, and $\theta_m \in [-\pi, \pi)$ denotes the angular direction of a trajectory at this point. Given specific coordinates in space and time: (x^*, y^*, t^*), we want the model to give us the probability distribution of the angular directions $p(\theta|x^*, y^*, t^*)$.

*Correspondence to: W. Zhi, weiming.zhi@sydney.edu.au
[1]School of Computer Science, The University of Sydney, Australia
[2]Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA
[3]NVIDIA Research, USA

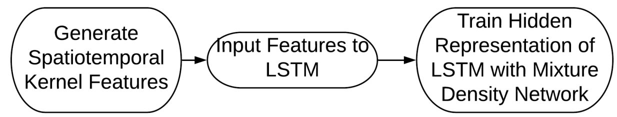

Fig. 1: Overview of our model

B. Overview of Proposed Model

A brief overview of our model is given in this subsection. Fig. 1 summarises the main architecture of our model. Given input data, we calculate high dimensional kernel features. The features are then treated as a sequence, and passed through a LSTM network. The hidden representation of the LSTM network is inputted to a mixture density network (MDN) to capture a multi-modal distribution. .

C. Generating High Dimensional Features Over Space

In this work, we approximate a full kernel matrix between each data point by projecting it to a set of pseudo-input points, similar to the sparse Gaussian Process method [4]. We start by defining a set of M_s pseudo-input points over space, denoted by $\bar{\mathbf{x}}_1 \dots \bar{\mathbf{x}}_{M_s}$. $\mathbf{x} \in \mathbb{R}^2$ and $\bar{\mathbf{x}} \in \mathbb{R}^2$, both $\mathbf{x}$ and $\bar{\mathbf{x}}$ contain longitudinal and latitudinal information. We generate high dimensional feature vectors from inputs by evaluating a kernel function, $k(\mathbf{x}, \bar{\mathbf{x}})$, between the input and each pseudo-input point.

The Radial Basis Function (RBF) kernel [5] is chosen to generate features in space. This kernel function is defined as $k(\mathbf{x}, \bar{\mathbf{x}}) = \exp\left(-\frac{\|\mathbf{x}-\bar{\mathbf{x}}\|_2^2}{2\ell^2}\right)$, where ℓ is the length scale hyperparameter of the kernel. Our kernel feature influences the neighbouring areas of training examples, resulting in the building of a continuous map.

D. From Spatial to Spatiotemporal

There are two steps to extend our model to spatiotemporal:

1) Generate high dimensional features in the space and time domains;
2) Pass the spatiotemporal features along the time domain into a LSTM network to learn a more compact representation of temporal variations

A natural way to build spatiotemporal kernels, described by [6], is to multiply a spatial kernel with a temporal one.

	Single Directional	Multi Directional
Uni-modal GM [3]	1.5	0.86
Multi-modal GM [3]	1.54	1.17
Continuous Map	1.69	1.44

TABLE I: The average likelihood values evaluated on benchmark datasets [3], using discrete methods from [3] and our continuous method

Similar to our method for spatial mapping, we assume M_t pseudo-input points in time, $\bar{t}_1, \ldots, \bar{t}_{M_t}$. Our spatiotemporal kernel function is given by

$$k\big((\mathbf{x}, t), (\bar{\mathbf{x}}, \bar{t})\big) = \underbrace{k_s(\mathbf{x}, \bar{\mathbf{x}})}_{\text{Spatial}} \underbrace{k_t(t, \bar{t})}_{\text{Temporal}}, \tag{1}$$

where $\mathbf{x}$ and t represent a coordinate in space and time, and $\bar{\mathbf{x}}$ and $\bar{t}$ are pseudo-input points in space and time respectively. In this work, we also use the RBF as the kernel function in time. We subsequently input the spatiotemporal features into a LSTM network [7]. The LSTM is able to learn a compact hidden representation for the spatiotemporal kernel features, which is a vector of the dimensionality of the specified output of the LSTM. A mixture density network is subsequently trained on the hidden representation of the LSTM.

E. Mixture Density Networks

We model the output of our model as a probability distribution of movement directions, with each mode in this distribution indicating a probable trajectory direction. We can achieve such outputs using mixture density networks (MDN) [8].

In a MDN, the probability density of the target distribution is represented as a convex combination of R individual probability density functions. We assume that the probability density function of angular directions of trajectories can be approximated by a mixture of von Mises distributions [9].

The loss function of the MDN is the average negative log-likelihood (ANLL) over all the training examples. We can then write the loss function as:

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N} \log\left(\sum_{r=1}^{R} \alpha_r \mathcal{VM}(\theta_n | \mu_m, \kappa_m)\right). \tag{2}$$

III. Experimental Results

We empirically demonstrate that the continuous nature of our model, provided by the kernel features, can capture changes in movement direction distribution space, as compared to a method that discretises the environment into independent cells, such as the methods presented in [3]. The average likelihood of the test examples using benchmark datasets used in experiments in [3] are presented in Table I. We also ran experiments to show the effect of temporal changes. Using the Benchmark Edinburgh Dataset [10], the directional distributions predicted at different times and different points in time are displayed Figure 2. We see that our model can capture the change of these distributions over time.

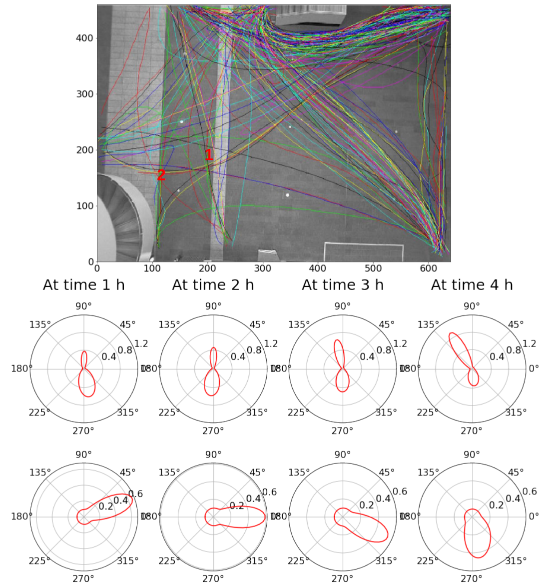

Fig. 2: The top figure shows trajectories people take. We attempt to model how the directional distribution change over time in two different locations marked as 1, 2. The predicted probability distributions over four time steps at the two locations is shown in the bottom image. The probability distributions are shown as polar plots, with the angular axis indicating the direction in degrees, and the radial axis specifying the probability density.

IV. Conclusion

By effectively exploiting the power of kernels and LSTMs to learn spatial and temporal patterns, we present a novel spatiotemporal model to learn the distribution of directions of moving objects. The proposed method is continuous in both space and time, and can capture the long-term dynamics in an environment. Future work shall look into exploiting periodic patterns in the directional distributions.

References

[1] D. Arbuckle, A. Howard, and M. Mataric, "Temporal occupancy grids: a method for classifying the spatio-temporal properties of the environment," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[2] T. Krajník, J. Pulido Fentanes, M. Hanheide, and T. Duckett, "Persistent localization and life-long mapping in changing environments using the frequency map enhancement," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[3] R. Senanayake and F. Ramos, "Directional grid maps: modeling multimodal angular uncertainty in dynamic environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018.

[4] E. Snelson and Z. Ghahramani, "Sparse gaussian processes using pseudo-inputs," in *Advances in Neural Information Processing Systems*, 2006.

[5] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. 2001.

[6] S. R. Flaxman, "Machine learning in space and time," tech. rep., Carnegie Mellon University, 2015.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, pp. 1735–1780, 1997.

[8] C. M. Bishop, "Mixture density networks," tech. rep., Aston University, 1994.

[9] K. V. Mardia, "Statistics of directional data," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 349–393, 1975.

[10] B. Majecka, "Statistical models of pedestrian behaviour in the forum," tech. rep., School of Informatics, University of Edinburgh, 2009.