Mixture of Switching Linear Dynamics to Discover Behavior Patterns in Object Tracks

Julian F. P. Kooij*
J.F.P.Kooij@tudelft.nl
Delft University of Technology

Gwenn Englebienne*
g.englebienne@utwente.nl
University of Twente

Dariu M. Gavrila
D.M.Gavrila@uva.nl
University of Amsterdam

Abstract

In [4] we present a novel non-parametric Bayesian model to jointly discover the dynamics of low-level actions and high-level behaviors of tracked objects. Actions capture both linear, low-level object dynamics, and an additional spatial distribution on where the dynamic occurs. Furthermore, behavior classes capture high-level temporal motion dependencies in Markov chains of actions, thus each learned behavior is a Switching Linear Dynamical System. The number of actions and behaviors is discovered from the data itself using Dirichlet Processes. We are especially interested in cases where tracks can exhibit large kinematic and spatial variations, e.g. person tracks in open environments, as found in the visual surveillance and intelligent vehicle domains. The model handles real-valued features directly, so no information is lost by quantizing measurements into 'visual words', and variations in standing, walking and running can be discovered without discrete thresholds. We describe inference using Markov Chain Monte Carlo sampling and validate our approach on several artificial and real-world pedestrian track datasets from the surveillance and intelligent vehicle domain. We show that our model can distinguish between relevant behavior patterns that an existing state-of-the-art hierarchical model for clustering and simpler model variants cannot. Software and various datasets are made public for benchmarking purposes.

1. Introduction

Probabilistic motion models are key to tracking [1], path prediction [5] in intelligent vehicles [2], and anomalous track detection in video surveillance [7]. Switching Linear Dynamical System (SLDS) [6] are suitable for maneuvering targets, such as walking and stopping pedestrians, but lack spatial/long-term temporal context on where/when motion states occur. On the other hand, Dual-HDP [7] discovers low-level spatially localized motion from track data, clusters these in high-level behavior classes, but relies on quantized position/motion features. This can lead to sparse data, even at low binning resolution, and is incompatible with the commonly used LDS. To obtain best of both, we propose (1) a hierarchical model to jointly discover what and how many low-level actions and high-level behavior classes are present in unlabeled track data, for which (2) we present an Markov Chain Monte Carlo (MCMC) inference scheme. Our model (3) extends SLDS switching states with spatial distributions (spatial context), and (4) can discriminate behaviors with different action orders (temporal context). (5) Features are not quantized, but actions capture motion and variance directly in the continuous feature space.

This paper is an abstract of our TPAMI publication [4].

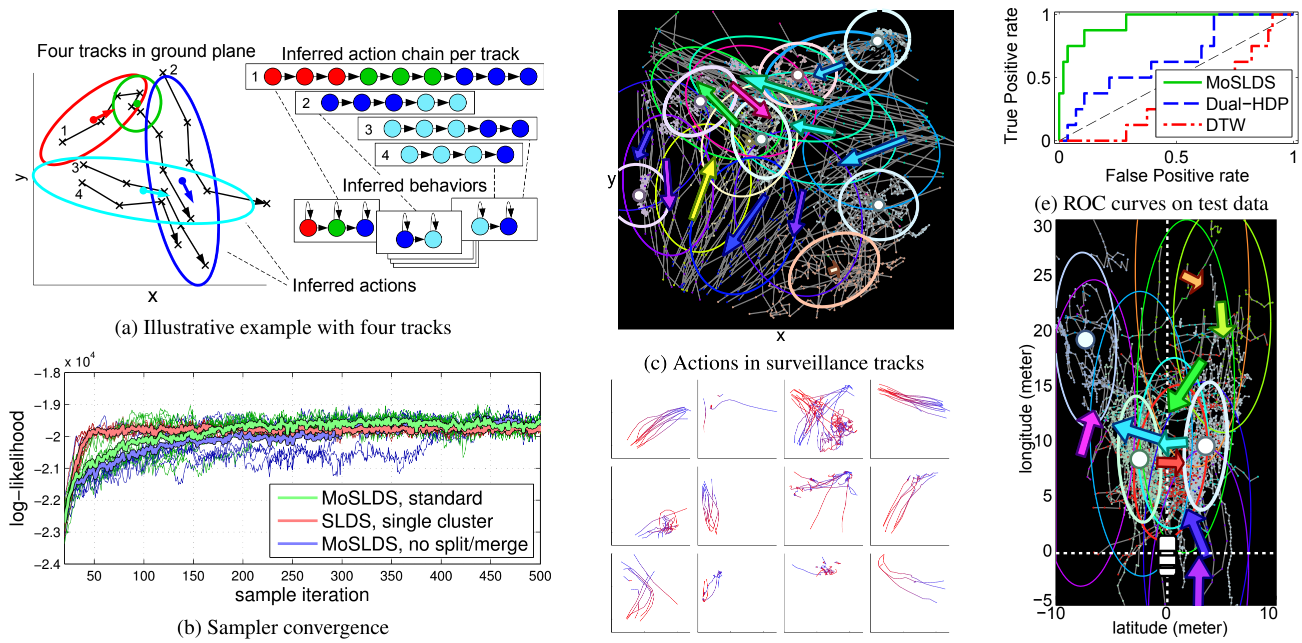
*The presented work was performed while all authors were affiliated with the University of Amsterdam, see [4].

2. Approach

In our approach, tracks are clustered into *behaviors*, each behavior defining transition probabilities between *action* states. Each action describes low-level motion dynamics as a LDS, and captures the spatial Gaussian distribution where the dynamic occurs. Fig. 1a illustrates how tracks are segmented into sequences of common actions, and jointly clustered into behavior classes with similar action sequences. Since each behavior is a SLDS, the full model is a Mixture of SLDS (MoSLDS). We describe a MCMC sampling scheme for Bayesian inference to perform action and behavior clustering jointly, which can improve learned dynamics. The number of actions and the number of behaviors are not fixed but discovered from the data itself using Dirichlet Process mixtures. Split-merge sampling steps can further help to avoid local optima and improve convergence. Sampling scales linearly with track count and length, but longer tracks may require more iterations to obtain good samples.

3. Experiments

We compare our MoSLDS to SLDS, Dual-HDP, and Dynamic Time Warping (DTW) on several artificial and real-world datasets from the surveillance and intelligent vehicle domains [5, 3]. For instance, the surveillance dataset con-



(a) Illustrative example with four tracks

(b) Sampler convergence

(c) Actions in surveillance tracks

(d) Behaviors in surveillance tracks

(e) ROC curves on test data

(f) Actions in pedestrian tracks

Figure 1: (a) MoSLDS example on four tracks (crosses are observed positions in ground plane). Observations are clustered into four actions (2D Gaussian regions with motion direction), and tracks into three inferred behavior clusters. (b) Convergence of various MCMC samplers. (c) Actions (as spatial region and mean LDS motion) in surveillance tracks (as gray lines), and (d) tracks clustered by behaviors. (e) ROC curves on test data. (f) Pedestrian actions around a vehicle (located at origin).

tains normative tracks of employees and visitors, but the test data also contains anomalies of people acting suspiciously or running away. The MoSLDS discovers actions for people entering/exiting and waiting, shown in Fig. 1c, and the behavior clusters that combine these, Fig. 1d. We find that Dual-HDP learns too specific actions for people standing still, and about half the test tracks contain quantized features not present in the training data. Thus, the ROC curves in Fig. 1e show that the MoSLDS is more capable of distinguishing the anomalies. Another dataset contains pedestrian tracks recorded with an embedded vision system in a vehicle. Fig. 1f shows that lateral motion is learned (small cyan & red arrows) of people crossing the road in front of the vehicle (located at the origin). Finally, Fig. 1b compares convergence of our full sampler, without split-merge jumps, and a single SLDS sampler over multiple MCMC runs.

4. Conclusions

Behavior clusters capture higher order dependencies on spatially localized dynamics, and provide insight in the behavioral structure of the data. Additional split-merge moves can help the sampler to avoid local optima more reliably, but on our real-world data all samplers eventually convergence to similar solutions. Unlike Dual-HDP, our MoSLDS does not rely on feature quantization. Learning dynamics and spatial context in the continuous feature space is beneficial when tracks have spatial and kinematic variations, such as people walking and standing in open spaces, and when available training data is limited.

References

[1] S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House Norwood, MA, 1999. 1

[2] C. G. Keller and D. M. Gavrila. Will the pedestrian cross? A study on pedestrian path prediction. *IEEE Trans. ITS* , 15(2):494–506, 2014. 1

[3] J. F. P. Kooij, G. Englebienne, and D. M. Gavrila. A non-parametric hierarchical model to discover behavior dynamics from tracks. In *Proc. ECCV*, pages 270–283. 2012. 1

[4] J. F. P. Kooij, G. Englebienne, and D. M. Gavrila. Mixture of switching linear dynamics to discover behavior patterns in object tracks. *IEEE PAMI*, PP(99):1–1, 2015. 1

[5] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proc. ICCV*, pages 261–268, 2009. 1

[6] A.-V. Rosti and M. J. F. Gales. Rao-Blackwellised Gibbs sampling for switching linear dynamical systems. In *Proc. of IEEE ICASSP*, volume 1, pages I–809, 2004. 1

[7] X. Wang, K. T. Ma, G. W. Ng, and W. E. Grimson. Trajectory analysis and semantic region modeling using a nonparametric Bayesian model. In *Proc. CVPR*, pages 1–8, 2008. 1