Unsupervised Trajectory Pattern Classification Using Hierarchical Dirichlet Process Mixture Hidden Markov Model

Conference Paper · September 2014

DOI: 10.1109/MLSP.2014.6958916

CITATIONS
20

READS
448

3 authors:

Vahid Bastani
Università degli Studi di Genova
13 PUBLICATIONS **238** CITATIONS
SEE PROFILE

Lucio Marcenaro
Università degli Studi di Genova
233 PUBLICATIONS **2,690** CITATIONS
SEE PROFILE

Carlo S. Regazzoni
Università degli Studi di Genova
580 PUBLICATIONS **7,216** CITATIONS
SEE PROFILE

UNSUPERVISED TRAJECTORY PATTERN CLASSIFICATION USING HIERARCHICAL DIRICHLET PROCESS MIXTURE HIDDEN MARKOV MODEL

Vahid Bastani, Lucio Marcenaro, Carlo Regazzoni

Department of Electrical, Electronic, Telecommunications Engineering
and Naval Architecture (DITEN), University of Genova, Genoa - Italy

ABSTRACT

In this paper we present a trajectory clustering method based on nonparametric Bayesian approach proposed for analyzing dynamic systems. Our method uses a modified hierarchical Dirichlet process-hidden Markov model in order to learn trajectory patterns into its parameter variables in an unsupervised way. Due to inherited Bayesian structure, this model resolves some limitations in trajectory clustering problem such as sequential analysis, incremental learning and non-uniform sampling. In this paper we introduce this model and its learning algorithm and finally we evaluate its performance.

Index Terms— Unsupervised Trajectory Clustering, Nonparametric Bayesian Learning, Motion Pattern Learning, Dirichlet Process Mixture

1. INTRODUCTION

Trajectory can be described as the history of changes in a dynamic system during the observation period of its state space. Therefore, in many fields that deal with dynamic systems it is desirable to be able to analyze and recognize different trajectory patterns. In the surveillance applications, activity of a subject is defined by its motion, which is represented by its trajectory. An intelligent surveillance system, in order to detect events and discriminate between different behaviors of subjects, should be able to classify and recognize trajectory patterns.

Trajectory learning is usually done by grouping similar trajectories into clusters because in real problems the motion pattern is a function of many complex processes. However, the problem is often divided in three sub-problems of trajectory representation, distance measure and clustering. Beside straightforward point-based representation of trajectories, some authors have used discrete Fourier transform coefficients, polynomial curve fitting and Haar wavelet transform to parameterize trajectories. The task of distance measure is to calculate the similarity between trajectories and serving this measure for clustering. For example, this measure can be computed using Euclidean point-base distance, principal component analysis, dynamic time warping, and longest common subsequence. Finally, clustering algorithms, which are used for grouping similar trajectories based on their relative distances, are generally divided into direct, agglomerative, divisive, hybrid, graph-based and spectral-based techniques. See [1] and [2] for more detail.

Proposed methods for trajectory learning has solved many application-specific problems but they have a number of known limitations:

- The concept of speed represented by relative time of observations usually disregarded in the proposed methods. However, in some applications, moving in a same path with different speed is considered as a completely different behavior.
- Distance measures usually require trajectories to have same length which is not the case in many applications. This issue specially is more important when we need to measure the distance sequentially as we receive new observations in a real-time system.
- With increasing interest in sparse signal representation, it is possible to have trajectories with non-uniform or random samples. However, current distance measures are not robust to randomness in sampling time.
- The majority of works so far are devoted to off-line, supervised and batch learning where there are available sufficient number of different trajectory patterns in the data set and they are appropriately labeled. Even when the data set is unlabeled and the method is unsupervised, the number of possible trajectory patterns should be known in advance.

Some works partly addressed above problems. On-line clustering has been studied in [3] and [4]. The method in [3] is also able to sequentially analysis the trajectories. Unsupervised learning is addressed in [5] and [2]. The later method

This work has been sponsored by Ministero della Istruzione, della Università e della Ricerca, DM62569 *C3 (Comando, Controllo, Comunicazioni) Tattico e ISR (Intelligence, Sorveglianza e Ricognizione)* Netcentrici con capacità di integrazione di elementi UXV C3ISR.

also devised for incremental learning. In the [6] the speed is also included in the trajectory feature vector.

To tackle above mentioned problems, in this paper we propose a unified probabilistic trajectory learning algorithm based on recent development in nonparametric Bayesian methods. Our method is able to cluster trajectories in unsupervised way and without knowing the number of different trajectory patterns. Learning can be done in off-line batch mode or on-line incremental mode. Furthermore, since distance measurement in our method is done in probabilistic way, when parameter of the model learned, we can compute likelihood that new trajectory belongs to each existing cluster or a new unseen cluster. In addition, trajectories are not needed to be same in length or have same sampling time. Finally, sequential analysis can be done even using a single point or the full data sequence.

The proposed method is based on an extension of Hierarchical Dirichlet Process-Hidden Markov Model (HDP-HMM). The Dirichlet process, $DP(\alpha, H)$ is a distribution on probability measures. The fact that the drawn measure from a DP is an infinite discrete measure motivates the use of this process as a prior to the distribution of mixture parameters in infinite mixture models. Such a model which called Dirichlet Process Mixture Models (DPMM) is efficiently able to tackle problems with unknown many number of cluster. Using DPMM concept, HDP-HMM is proposed as a HMM with infinite or unknown number of hidden states. The successful performance of HDP and HDP-HMM has been shown in [7] and [8] for global scene understanding where no prior information is available.

In the proposed method trajectory clusters are modeled as a series of transitions in the hidden state of HMMs. Each cluster has its own transition pattern, thus it is a different HMM. However, the hidden states which have a one-to-one relation to a set of spatio-temporal observation clusters (STC) are shared among all HMMs. The HMM and DPMM methods separately have been used in the same context before. In the recent work of [2] authors have employed DPMM for incremental clustering of unknown number of trajectory pattern. However, clustering is done in the Fourier feature space where time is not taken into account. Conventional HMM has been used for this problem long before, in the early work of [9], it is used for trajectory distance measure and recently in [10], it is used for analysis the motion pattern of crowed.

2. TRAJECTORY REPRESENTATION AND SPATIO-TEMPORAL POINT CLUSTERS

A Trajectory j is represented by time series $\boldsymbol{\chi}_j = \{\mathbf{x}_1^j, \mathbf{x}_2^j, \cdots, \mathbf{x}_{N_j}^j\}$, where N_j is the length of the trajectory. In this work we assumed that each point $\mathbf{x}_i^j = [x_i^j, y_i^j, t_i^j]^T$ is a vector in in three dimensional time-position space where x_i^j and y_i^j correspond to position and t_i^j corresponds to time. Furthermore, we assumed that the observation time is relative to the time of the first observation such that $t_1^j = 0$. Relative time is used in the observation vector in order to make it possible to discriminate between same trajectories with different speed.

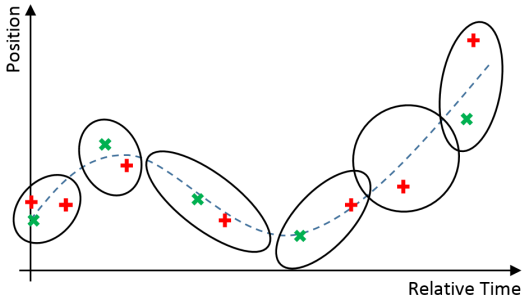

Fig. 1. Example of trajectory representation and spatio-temporal clusters (STCs). Dashed line is the actual trajectory, red plus and green cross markers are two randomly sampled and noise affected instant of the actual trajectory, Ovals are STCs correspond to trajectory sections.

We first assume that each point $\mathbf{x}_i^j$ belongs to a trajectory section we call STC. Then we assume that points in a STC have a parametric probability distribution function (PDF) with parameter θ. In this way, we can easily compute the distance of a point to an STC using the value of the likelihood of that point being generated by corresponding PDF. In addition, since we have a probabilistic representation of the time of the observations, it is possible to handle sparse and random sampling. Fig.1 shows an example of how STCs work, dashed line shows the actual trajectory in relative time and position (note that time is strictly increasing). Red plus and green cross markers are depicting two time series instantiated from the same actual trajectory but sampling times are random and positions are subjected to noise. Ovals show STCs that are trying to group random observations in order to model sections of the trajectory.

In this work, we chose three variate Gaussian distribution as the PDF of STCs. Thus the PDF parameters are $\theta = \{\Sigma, \mu\}$, namely the covariance matrix and mean vector of the Gaussian. In the mixture model analysis it is usually desirable to learn component parameters as a posterior of another parametric distribution which is conjugate to the distribution of the observations. This conjugacy simplifies Bayesian inference of parameters since the posterior distribution has same form as the conjugate prior. For multivariate Gaussian, the conjugate prior is Normal Inverse Wishart (NIW) distribution which has four parameters $\lambda = \{\mu_0, \kappa, \Psi, \nu\}$ which can be fixed as a non informative prior or learned from data, see [11] for more detail.

3. INFINITE MIXTURE OF INFINITE STATE HMM

HMM is one of simplest form of dynamic Bayesian networks (DBN) for modeling discrete state-space dynamic systems. However, more complex DBNs are also used in [12] and [13] for almost the same application. Following Markov process assertion, in HMM model it is assumed that given latent state variable of the system, observations are stochastically independent. Although this is a very restrictive assumption, it makes analysis of complex processes significantly tractable. However, in the ordinary HMM the cardinality of state-space should be known in advance which is not always possible for real systems with many unspecified modes.

Suppose that we have J set of observations as time series $\boldsymbol{\chi}_j$ and each time series represents a trajectory j. Our aim is to group similar observations into STCs, then capture the dynamic pattern of the trajectory using a HMM on these STCs. We assume that there is no prior information about the number and the size of clusters and we want to learn these from data. More importantly, we assume that there are many different HMMs of this kind with shared STCs such that each HMM encodes a different behavior (trajectory) in its own state transition matrix. In each resulting individual HMM, the required number of hidden states and the parameters of the emission distributions (probability distribution function of an observation given the state variable) is not known a priori and each individual HMM should have a significantly different transition matrix from others.

Using the capability of DPMM to tackle problems of unknown many or infinite mixture distribution and the idea of hierarchical chaining of these processes, also inspired by [14], HDP-HMM is proposed in [15]. It is an extension of finite state HMM to a model that is potentially able to deal with infinite number of states and emission distribution. However, HDP-HMM necessarily results a single HMM with its specific possibly infinite-dimensional transition matrix. In order to model undefined number of different HMMs, We extend the HDP-HMM by introducing another top level DPMM on the hierarchy to make a model of infinite mixture of HDP-HMMs. In order to explain the idea, first we establish the infinite mixture of finite state HMMs then we extend it to infinite mixture of HDP-HMMs. Suppose each observation sequence $\boldsymbol{\chi}_j$ is drown from an infinite mixture of HMMs with K possible hidden states as follows:

$$\boldsymbol{\chi}_j \sim \sum_{l=0}^{\infty} \rho_l \mathcal{H}(N_j; \Pi_l, \Theta_l). \tag{1}$$

where ρ_l is the mixture weight of the lth component and $\mathcal{H}(N_j; \Pi_l, \Theta_l)$ is a distribution of a sequence of N_j observation according to a HMM model with $K \times K$ state transition matrix Π_l and the set of parameters of emission distributions $\Theta_l = \{\theta_{l,1}, \cdots, \theta_{l,K}\}$. We let the emission parameters be shared (fixed) among mixture component $\forall l : \Theta_l = \Theta$ where $\Theta = \{\theta_1, \cdots, \theta_K\}$. Each row k of transition matrix Π_l is a categorical conditional probability distribution $\boldsymbol{\pi}_{k,l}$ of the next state given current state being k in the lth HMM. Since we want to encode dynamic behavior in the transition matrices each $\boldsymbol{\pi}_{k,l}$ should be distinct with respect to the current state k and the component l. In other words, we want different HMMs and their different states not to be able to contribute to update the parameters of the state transition PDF of each other. This requirement prevents us to share Π_l parameters among components as it is done in the case of Θ_l.

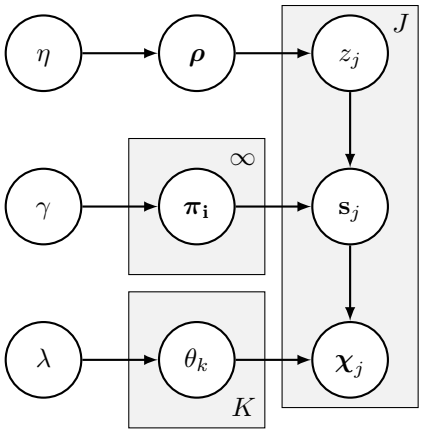

Fig. 2. Dirichlet Process Mixture of HMMs from Eq.1

Using DPMM model the process of Eq.1 can be represented as a Bayesian network of Fig.2, where $\mathbf{s}_j = \{s_1^j, s_2^j, \cdots, s_{N_j}^j\}$ is the sequence of hidden variables. The index of $\boldsymbol{\pi}_{\mathbf{i}}$ can be thought as a vector $\mathbf{i} = [l, k]_{k=1:K}^{l=1:\infty}$ to distinguish between different combinations of HMMs and their states. Given prior hyper-parameters γ, η and λ we can write conditional probability distributions of this graph as follows. We use $.|.$ to show condition,

$$\theta_k \sim H(\lambda). \tag{2}$$

$$\mathbf{x}_n^j | s_n^j, \theta_k \sim F(\theta_{s_n^j}). \tag{3}$$

Generally F and H can be any pair of conjugate distributions but in our case they are Gaussian and NIW respectively. In Fig.2, z_j is an indicator variable that shows the index of HMM for each time series j and it has conditional distribution as follows:

$$\boldsymbol{\rho} \sim GEM(\eta). \tag{4}$$

$$z_j | \boldsymbol{\rho} \sim Cat(\boldsymbol{\rho}). \tag{5}$$

Cat means categorical (multinomial) distribution. $\boldsymbol{\rho}$ is HMMs mixture weights which is a draw from Griffiths-Engen-McCloskey (GEM) stick braking process. The samples of this distribution can be used as parameter for infinite categorical distribution. Finally, the state transition process is as follows:

$$s_{n+1}^j | s_n^j, z_j, \boldsymbol{\pi}_{\mathbf{i}} \sim Cat(\boldsymbol{\pi}_{\mathbf{i}^*}), \text{Where } \mathbf{i}^* = [z_j, s_n^j] \tag{6}$$

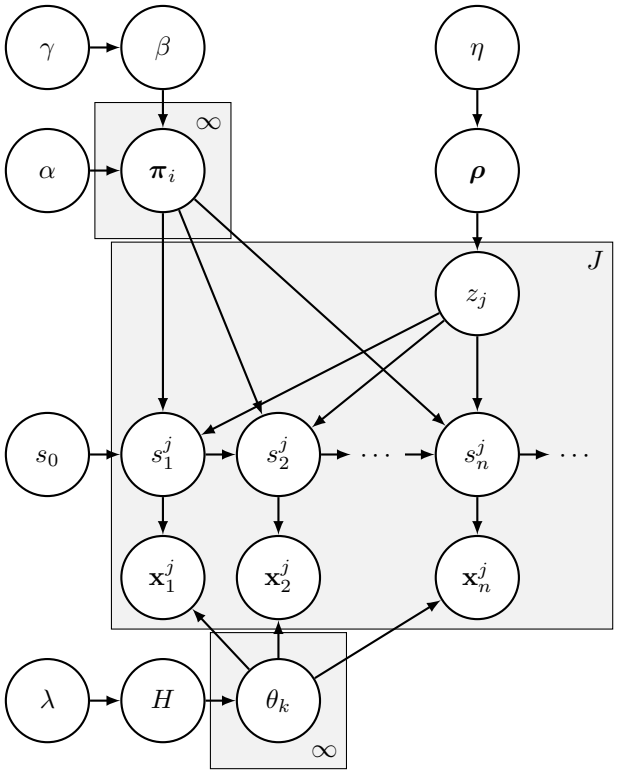

Fig. 3. Generative probabilistic graphical model of trajectory patterns using infinite mixture of HDP-HMMs

So far HMMs in our model has finite number of states which implies the number of STCs to be known. However, following the approach given by [15], a HMM with countably infinite number of states can be derived from K state HMM mixture of Equations (1) to (6) by letting $K \rightarrow \infty$. This model is depicted in Fig. 3, where, unlike Fig. 2, the state time series and observation time series is shown in the expanded view. The difference is that in the HDP-HMM state transition PDF is drawn from a Dirichlet process with concentration parameter α and base measure β which itself has a GEM distribution with parameter γ:

$$\beta \sim GEM(\gamma). \tag{7}$$

$$\pi_i \sim DP(\alpha, \beta). \tag{8}$$

Since the model in Fig.3 is a Bayesian network, many of the clustering problems reduce to Bayesian inference. For instance, incremental learning and sequential analysis can be done using well known recursive Bayesian update equation $p(\Gamma, x_{1:n+1}^j) = p(\Gamma, x_{1:n}^j)p(x_{n+1}^j|\Gamma)$, where Γ is the set of all model parameters. In addition, regardless of sampling time and the length of the trajectories, classification is simplified by Bayesian maximum likelihood and maximum a posterior probability classifier.

4. LEARNING AND INFERENCE

The Gibbs sampling technique is usually used to perform learning and inference in DPMM models. Gibbs sampling is a variant of Markov Chain Monte Carlo technique which iteratively sample every variable in the model conditioning on that other variables are fixed. Iterations proceed until all variables relax on equilibrium. Different variation of this algorithm has been proposed for HDP-HMM model such as collapsed, direct, block and weak limit Gibbs sampler, interested readers may refer to [16] for more detail. In this work, we used straightforward Gibbs sampler. In the following we briefly introduce sampling equations in order of execution.

1. **Sampling Hidden States** $\mathbf{s}_j$: for each sequence j sample hidden states $\mathbf{s}_j = \{s_1^j, s_2^j, \cdots, s_{N_j}^j\}$ sequentially, by integrating out z_j with following equation.

$$\begin{aligned} p(s_n^j|\mathbf{x}_{1:n}^j, \theta_k) &= \sum_{z_j} p(s_t^j|\mathbf{x}_{1:n}^j, \theta_k, z_j) \\ &\propto \sum_{z_j} p(s_n^j, \mathbf{x}_n^j|s_{1:n-1}^j, \mathbf{x}_{1:n-1}^j, \theta_k, z_j) \\ &\propto p(\mathbf{x}_n^j|s_n^j, \theta_k) \sum_{z_j} \sum_{s_{n-1}^j} p(s_n^j|s_{n-1}^j, z_j) \\ &\qquad p(s_{n-1}^j|\mathbf{x}_{1:n-1}^j, \theta_k, z_j) \end{aligned} \tag{9}$$

2. **Sampling HMM Indicator Variable** z_j: for each sequence j sample z_j given its hidden state sequence.

$$\begin{aligned} p(z_j|s_{1:N_j}^j, \boldsymbol{\rho}, \boldsymbol{\pi}_i) &\propto p(s_{1:N_j}^j|z_j, \boldsymbol{\pi}_i)p(z_j|\boldsymbol{\rho}) \\ &\propto p(z_j|\boldsymbol{\rho}) \prod_n p(s_n^j|s_{n-1}^j, z_j, \boldsymbol{\pi}_i) \end{aligned} \tag{10}$$

3. **Sampling HMM Mixture Weights** $\boldsymbol{\rho}$: let N_l be the number of sequences assigned to lth HMM, then based on Chinese restaurant process we can sample $\boldsymbol{\rho}$ given its prior η [15].

$$p(\boldsymbol{\rho}|z_{1:J}) = Dir(N_1 + \eta, \cdots, N_L + \eta, \eta) \tag{11}$$

4. **Sampling HMM Transition Probabilities** $\boldsymbol{\pi}_i$: let $N_{k,k',l}$ be the number of times a transition occurred from state k to state k' in the lth HMM, then the conditional distribution of $\pi_{i^*=[k,l]}$ is as follows [15].

$$\begin{aligned} &p(\boldsymbol{\pi}_{i^*=[k,l]}|\mathbf{s}_{1:J}, z_{1:J}, \alpha, \beta) = \\ &Dir(N_{k,0,l} + \alpha\beta_0, \cdots, N_{k,K,l} + \alpha\beta_K, \alpha \sum_{i=K+1}^{\infty} \beta_i) \end{aligned} \tag{12}$$

5. **Sampling DP base distribution** β: we first sample an auxiliary variable $M_{k,k'}$ from following conditional distribution

$$p(M_{k,k'} = M|\mathbf{s}_{1:J}, \beta, \alpha) \propto S(N_{k,k'}, M)(\alpha\beta'_k)^m \quad (13)$$

where $S(.,.)$ is Stirling number of the first kind, and $N_{k,k'} = \sum_l N_{k,k',l}$. Then the DP base distribution is sampled by

$$p(\beta|M_{0:K,0:K}, \gamma) = Dir(M_{.0}, \cdots, M_{.K}, \gamma) \quad (14)$$

where $M_{.k'} = \sum_k M_{k,k'}$ [15].

6. **Sampling cluster parameters** θ_k: if $H(.)$ chosen to be the conjugate prior of $F(.)$, then posterior probability distribution of each θ_k will be:

$$p(\theta_k|\mathbf{s}_{1:J}, \boldsymbol{\chi}_{1:J}, H) = H(\lambda^*) \quad (15)$$

where λ^* is the updated hyper parameter given current $\boldsymbol{\chi}_{1:J}$ and $\mathbf{s}_{1:J}$.

For trajectory classification purpose, one may use Equations (9) and (10) for iterative sampling while other parameters are kept unchanged. The resulting sample value that occurs most commonly is essentially equivalent to maximum a posteriori estimation of HMM indicator variable z_j and hidden states (STCs) sequence $\mathbf{s}_j$.

5. EXPERIMENT

For evaluating the learning algorithm, three different trajectory patterns with high sampling rate have been generated synthetically. Then, for each trajectory pattern ten independent sequences have been generated by first down-sampling the trajectory using random time instances and then adding independent Gaussian noise to the position of samples. Random sampling is conducted by first choosing a fixed sampling interval and then for each sample, adding random variable to the sampling time. In this experiment this random variable is a zero mean uniformly distributed over 30% bound of sampling interval. An example of resulting data is depicted in Fig.4 in three-dimensional position and relative-time space. The same procedure described in [12] for maritime vessels behavior simulation in port areas is employed here for data generation. However, random sampling used here can be thought as random time of receiving position information from AIS tracking system.

The hyper-parameters γ, α and η has been set uninformatively. However it is important to choose an small value for α in order to get more divers forms of state transition distributions. The hyper-parameter λ which is the set of the parameters of the NIW distribution has been chosen such that Gaussian mean and covariance matrix draws from this distribution fall more often in the observation region. However, these parameters can also be learned from data.

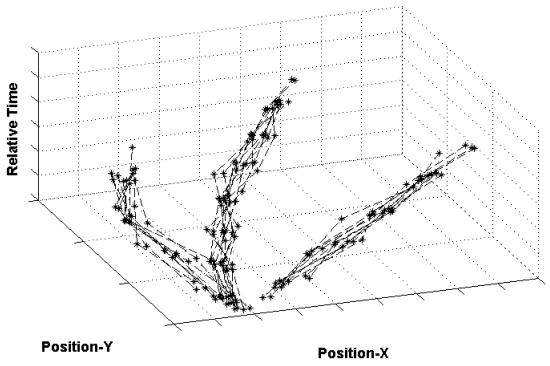

Fig. 4. Example of the data from three trajectory patterns used to evaluate learning algorithm.

We used adjusted mutual information (AMI) metric [17] for clustering performance assessment. Given the ground-truth of the cluster assignments this metric compares the result of clustering by assigning values in range [0 1] to it. Values close to zero means two assignment are highly independent and values close to one indicate agreement in them. The effect of random match and different permutation is eliminated in this measure.

The Gibbs sampling algorithm of Section 4 has been run over the synthetic trajectories and for each iteration the AMI of clustering assignment has been recorded. The experiment has been conducted 30 times with independently generated data. The mean and the standard deviation region of the AMI between clustering assignment and the ground-truth for each iteration is depicted in Fig.5. As can be seen from Fig.5, the Gibbs sampler converges relatively fast in the first 20 to 40 iterations. However, as it is known for the Gibbs sampling [18], sometimes it get trapped in some local optimum variables for long time (at least the first 200 iterations shown here). For this reason the mean AMI is at most about 0.8 in this experiment. More sophisticated Gibbs sampler may be used in future to tackle this problem.

6. CONCLUSION AND FUTURE WORKS

This paper described a new nonparametric Bayesian approach for trajectory clustering problem. The method used here is a modification of hierarchical Dirichlet process-hidden Markov model in which a new hierarchy level is introduced in order to achieve an infinite mixture of infinite state hidden Markov model. The basic idea of the model is to cluster point of the trajectory into some spatio-temporal clusters then learn the trajectory dynamic patterns into the transition probabilities of these clusters.

the resulting model is a Bayesian network and it can resolve many issues in trajectory clustering problem such as incremental learning and sequential analysis.

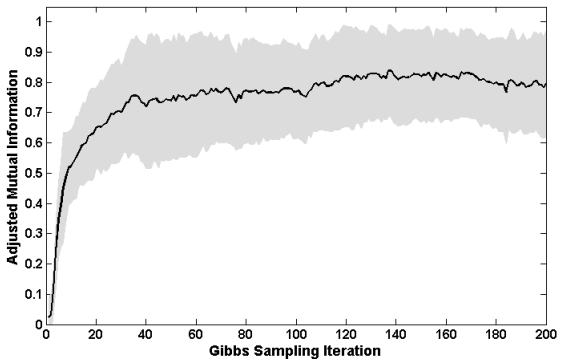

Fig. 5. AMI of cluster assignment for each iteration of Gibbs sampler. The solid line is the mean and the gray region demonstrate standard deviation bound.

The experimental evaluation shows that the model is able to learn trajectory patterns in unsupervised way. However, some problems still exist with convergence rate of the Gibbs sampling learning algorithm. An important future work direction is improving mixing rate of this learning algorithm using more sophisticated Gibbs sampler or revision in the model.

7. REFERENCES

[1] B. Morris and M. Trivedi, "Learning trajectory patterns by clustering: Experimental studies and comparative evaluation," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, June 2009, pp. 312–319.

[2] W. Hu, X. Li, G. Tian, S. Maybank, and Z. Zhang, "An incremental DPMM-based method for trajectory clustering, modeling, and retrieval.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 5, pp. 1051–65, May 2013.

[3] G. Laxhammar, R.and Falkman, "Online Learning and Sequential Anomaly Detection in Trajectories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2013.

[4] C. Piciarelli and G.L. Foresti, "On-line trajectory clustering for anomalous events detection," *Pattern Recognition Letters*, vol. 27, no. 15, pp. 1835–1842, 2006.

[5] H. Liu and J. Li, "Unsupervised multi-target trajectory detection, learning and analysis in complicated environments," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, Nov 2012, pp. 3716–3720.

[6] L. Song, F. Jiang, Z. Shi, R. Molina, and A. K. Katsaggelos, "Toward Dynamic Scene Understanding by Hierarchical Motion Pattern Mining," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–13, 2014.

[7] R. Emonet, J. Varadarajan, and J. Odobez, "Temporal analysis of motif mixtures using Dirichlet processes.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 1, pp. 140–56, Jan. 2014.

[8] D. Kuettel, M.D. Breitenstein, L. Van Gool, and V. Ferrari, "What's going on? discovering spatio-temporal dependencies in dynamic scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 1951–1958.

[9] F. Porikli, "Trajectory distance metric using hidden markov model based representation," in *IEEE European Conference on Computer Vision, PETS Workshop*, 2004.

[10] M. Andersson, F. Gustafsson, L. St-Laurent, and D. Prevost, "Recognition of Anomalous Motion Patterns in Urban Surveillance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 102–110, 2013.

[11] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis, Second Edition*, Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2003.

[12] F. Castaldo, V. Bastani, L. Marcenaro, F. Palmieri, and C. Regazzoni, "Abnormal Vessel Behavior Detection in Port Areas Based on Dynamic Bayesian Networks," in *17th IEEE International Conference on Information Fusion*, 2014 In Press.

[13] A. Dore and C. Regazzoni, "Interaction Analysis with a Bayesian Trajectory Model," *IEEE Intelligent Systems*, vol. 25, no. 3, pp. 32–40, May 2010.

[14] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden markov model," in *Machine Learning*. 2002, pp. 29–245, MIT Press.

[15] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. Blei, "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

[16] E. B. Fox, *Bayesian nonparametric learning of complex dynamical phenomena*, Ph.D. thesis, MIT, 2009.

[17] N. X. Vinh, J. Epps, and J. Bailey, "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance," *The Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, Mar. 2010.

[18] E. B. Sudderth, *Graphical Models for Visual Object Recognition and Tracking*, Ph.D. thesis, MIT, 2006.