Bayesian Inference Application: Traffic Pattern Analysis

Jin Young Choi

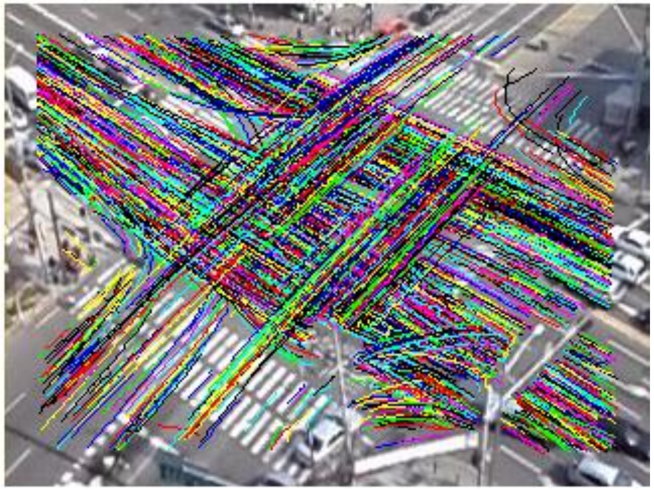
Traffic Pattern Analysis

- Surveillance in crowded scenes



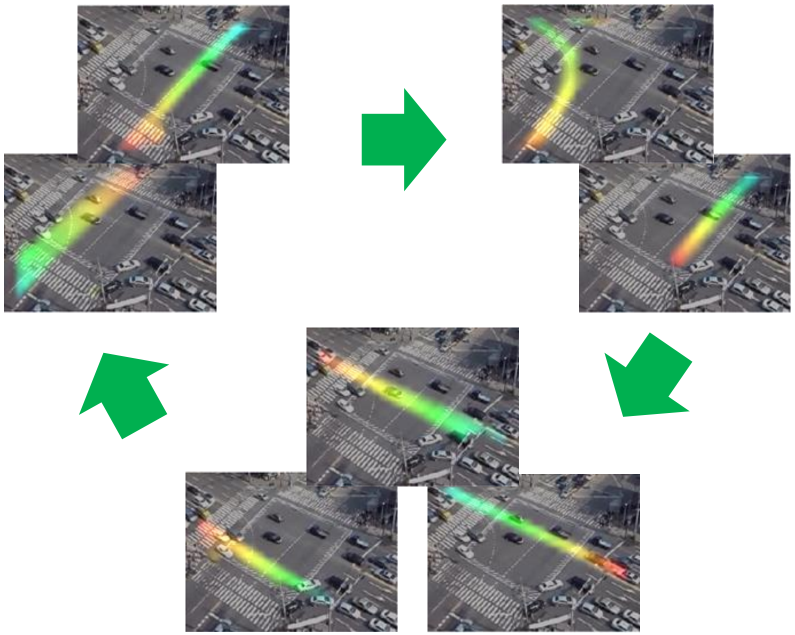

Problem Statements

- Unsupervised Traffic Pattern Analysis



Problem Statements

- Anomaly Detection using the Traffic Patterns
- Video summarization
- Prior motion model for tracking



RQ1: Semantic regions of normal pattern

- Clustering & Combining motion segments
- Optical flows

- Broken trajectories

RQ2: Speed information

- Discrimination ability of speed difference:
  - Bikes running in pedestrian road
  - Cars driving with over speed
  - Cars stopping in a railroad crossing
  - Pedestrians walking along the path of vehicles

RQ3: Interaction of trajectory patterns

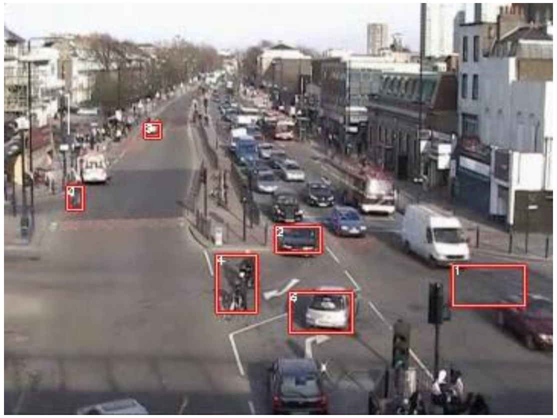
- Interaction of trajectory patterns
  - Activity modeling often involves modeling of interactions (between cars, people and environment)
  - Some activity can be normal or abnormal based on other co-occurring activities

normal

Abnormal (conflict!)

RQ4: Robust to crowded scene

- In crowded scenes, it is hard to extract motions of individual objects
- Crowded scenes cause frequent tracking failures, producing many broken trajectories

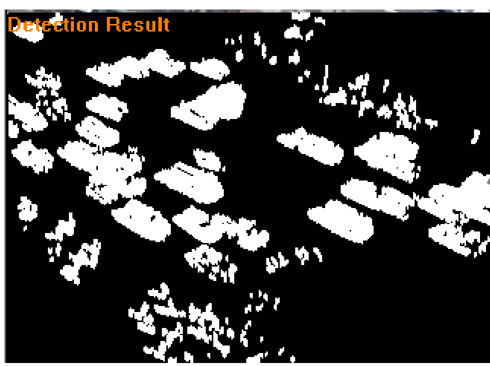


RQ5: Online Adaptation

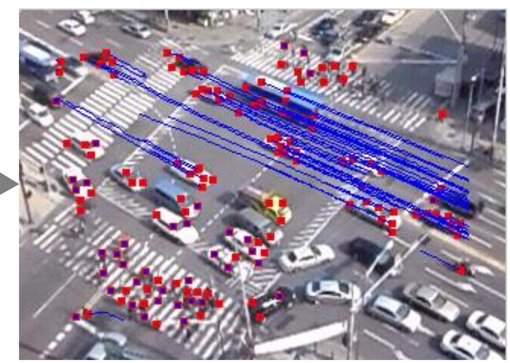
- The model should be able to adapt itself to temporal changes of the scene (e.g. reversible lane, traffic volume changes).
- It can also save memory and computational load since the model does not need to keep old data.

Motion and Trajectory Extraction

- Corner Point Detection on Foreground
- KLT tracking

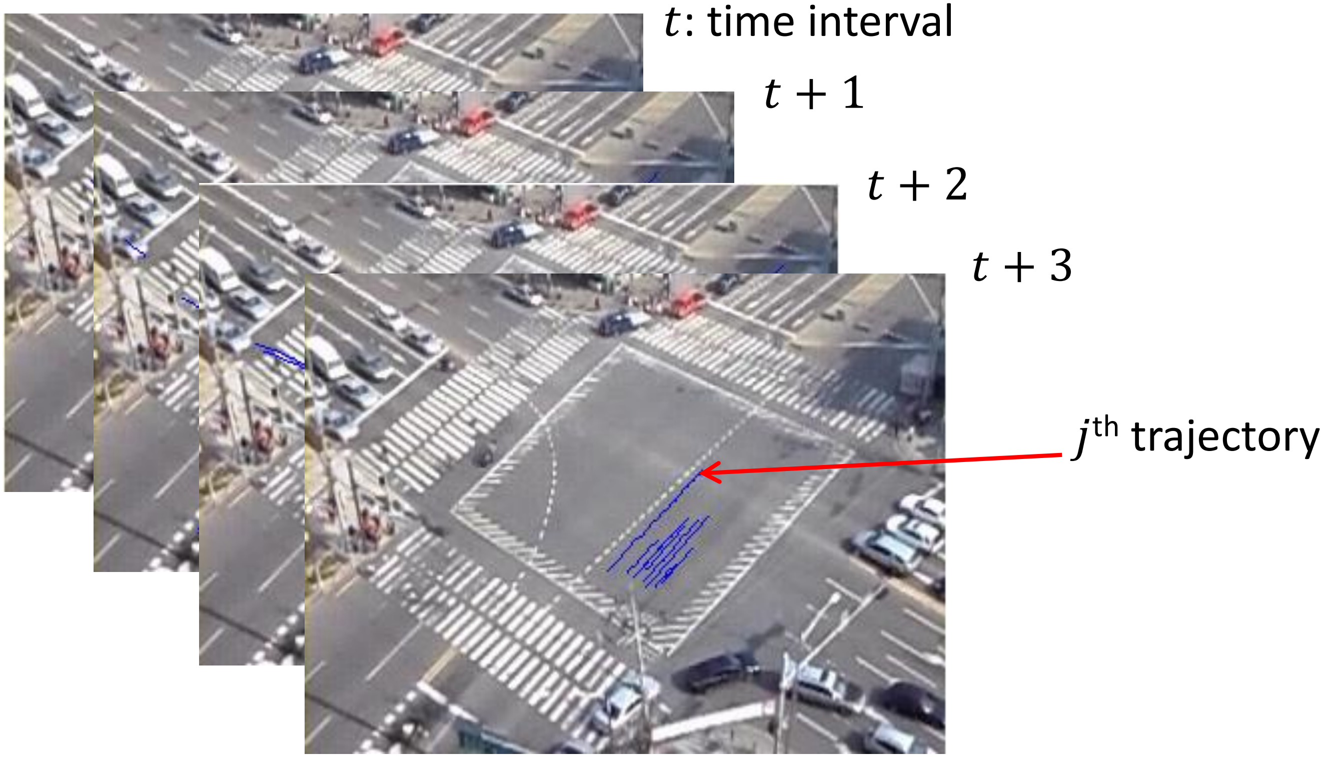


VS.



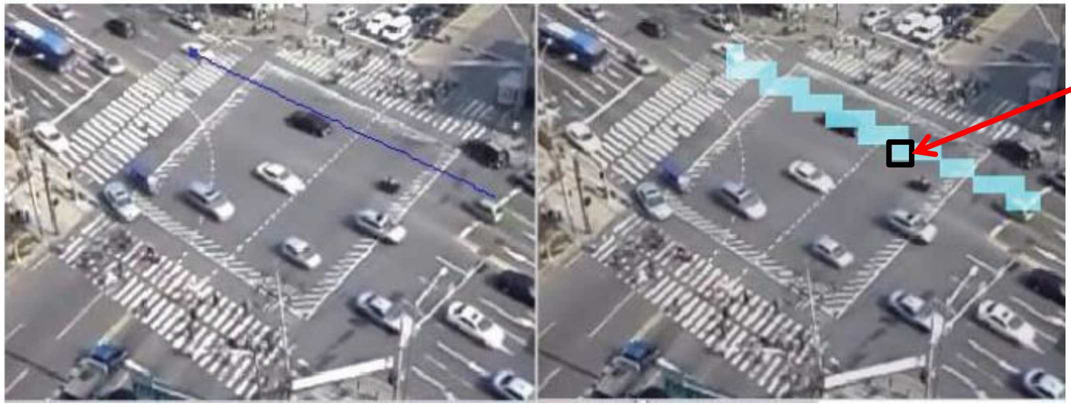
Input Data: Trajectories

- Trajectory collections



Trajectory representation: Location

- Coarse-level (quantization)
- Trajectory Patterns: modelled as Topic
- Passing Regions(cells): modelled as Words

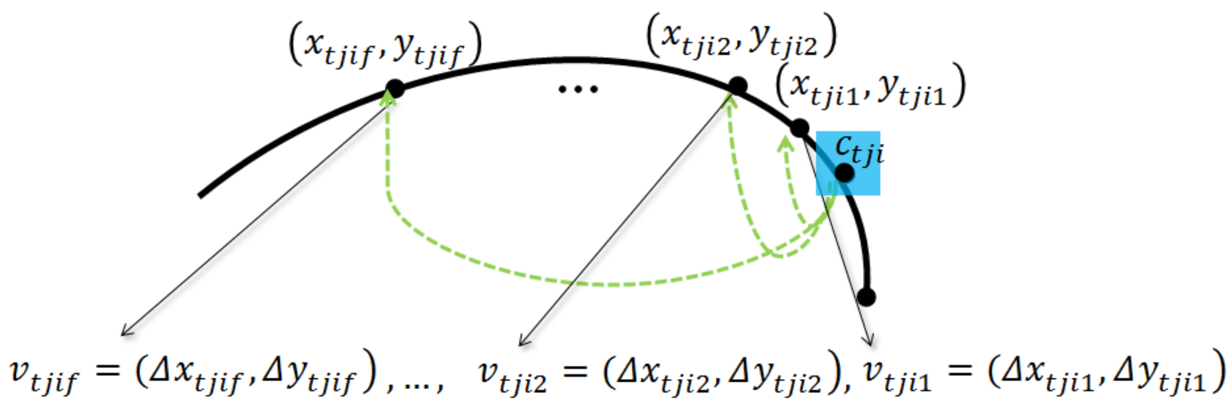


c_{tji}

t: time interval
j: trajectory index
i: cell index

Trajectory representation: Speed

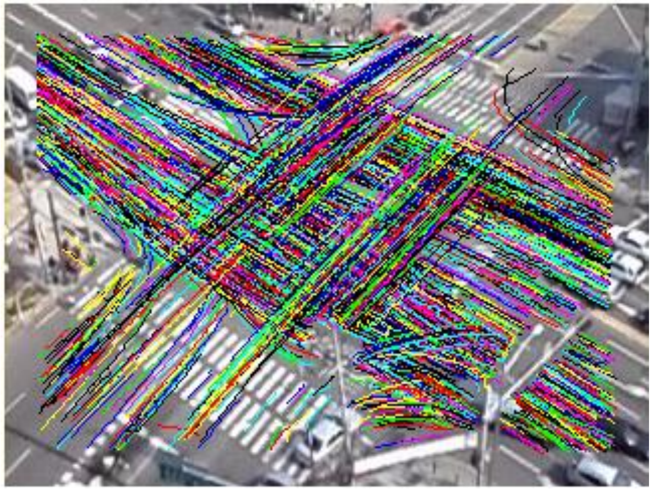
- Fine-level
  - Average velocity for f-frames



Generative model

- Model learning w.r.t observations

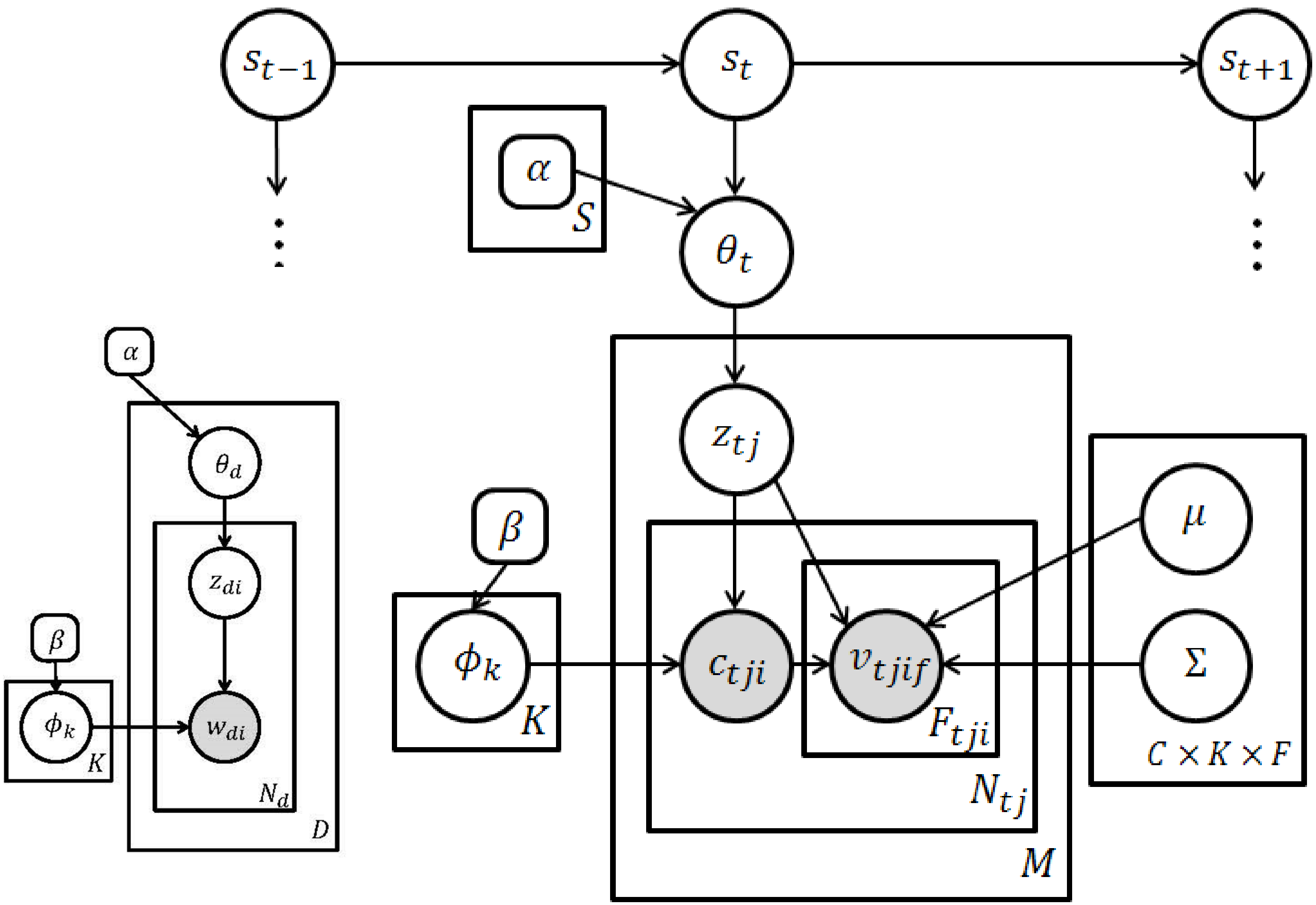
$$\underset{\mathcal{H}}{\text{argmax}}\ p(\mathcal{H}|\mathcal{O})$$



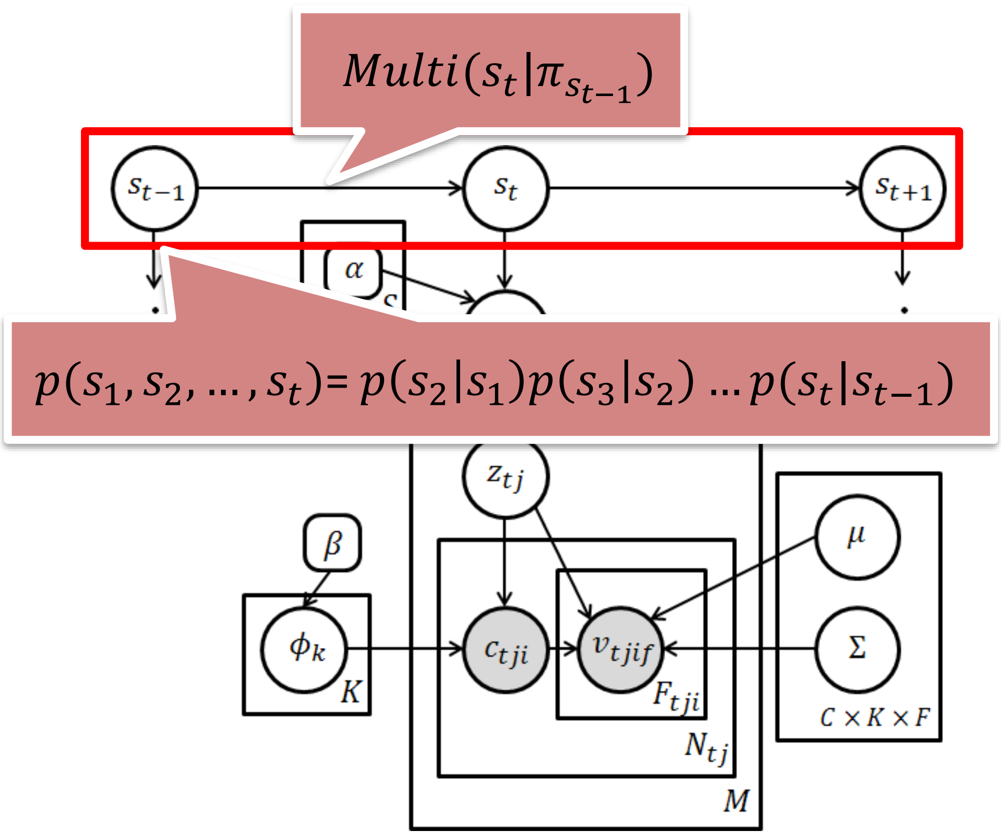
$$\mathcal{O} = \{c_{tji}, v_{tjif} | \forall t, i, j, f\}$$

- Test newly observed trajectories using the trained model

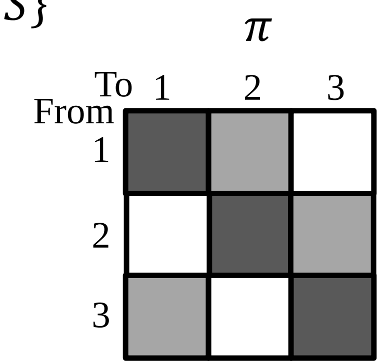
Graphical Inference Model

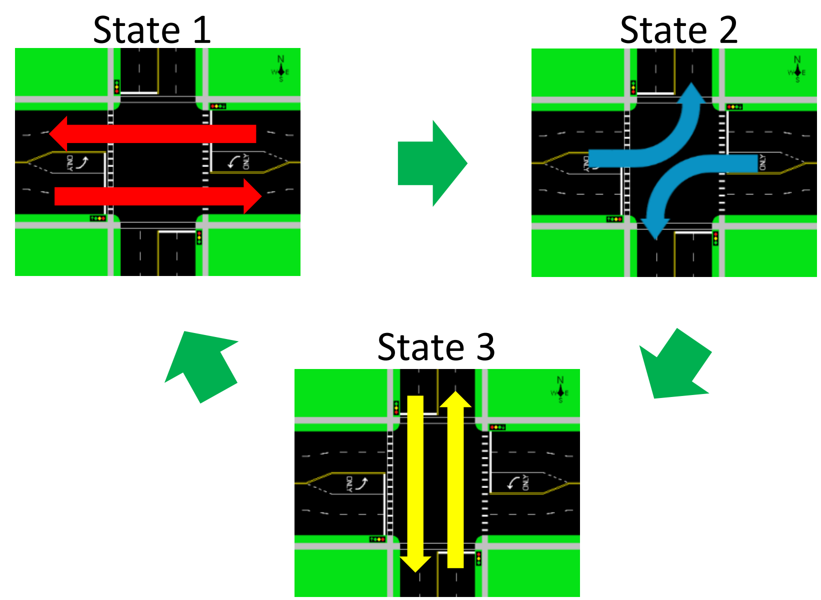


State Transition



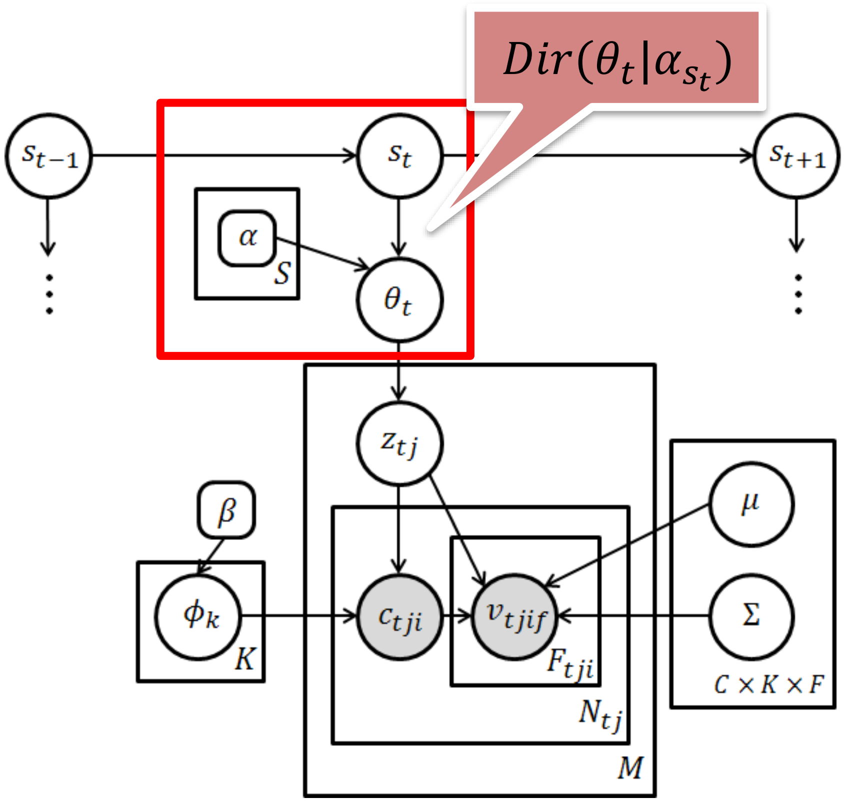
$$s_t \in \{1, \ldots, S\}$$

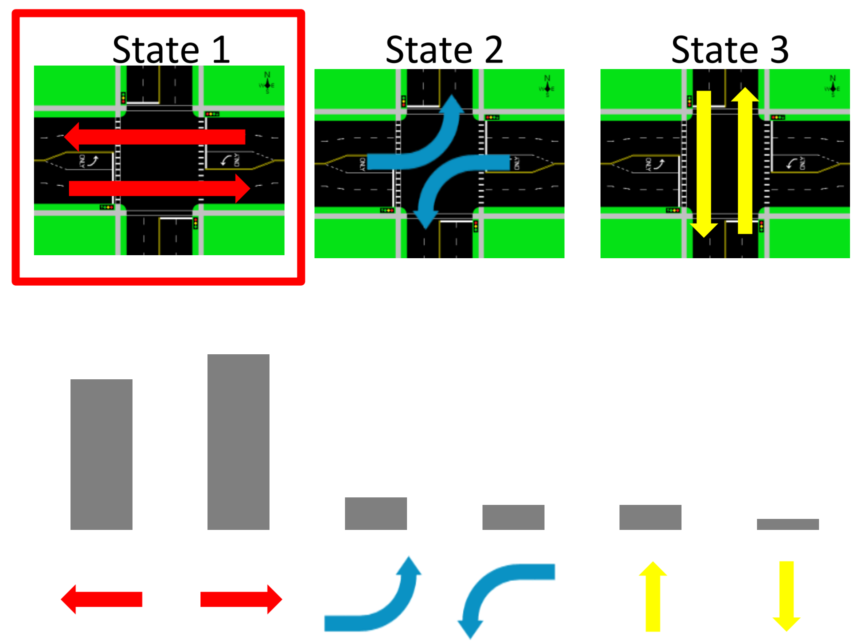




Distribution of Topic Occurrence

- Topic proportion

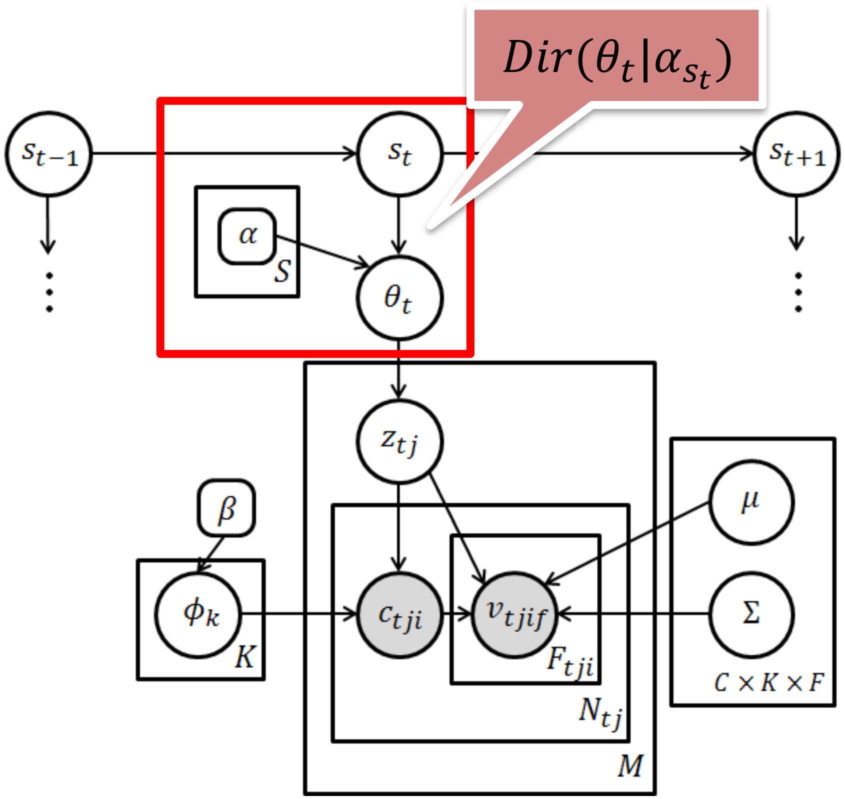


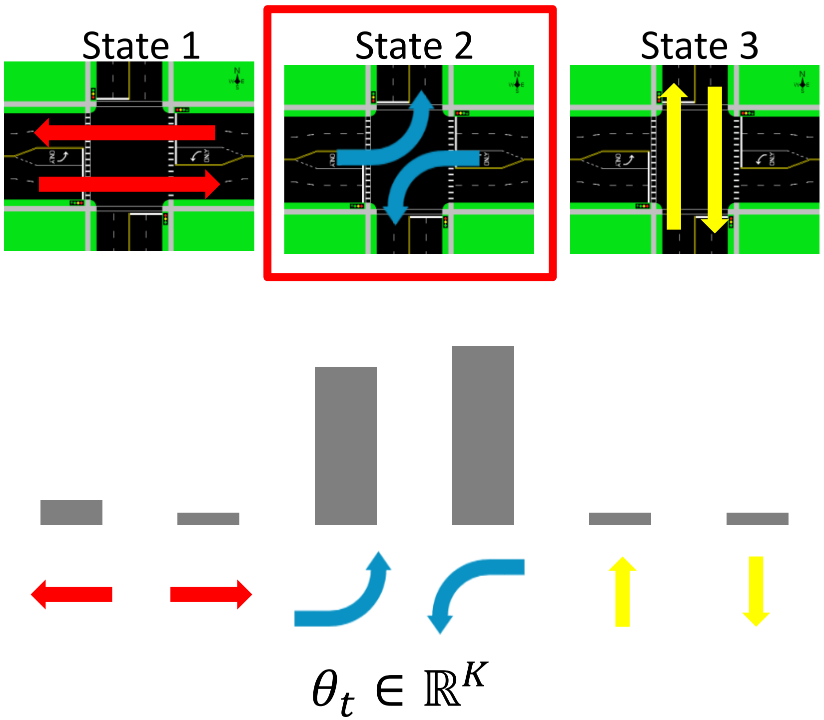


$\theta_t \in \mathbb{R}^K$

Distribution of Topic Occurrence

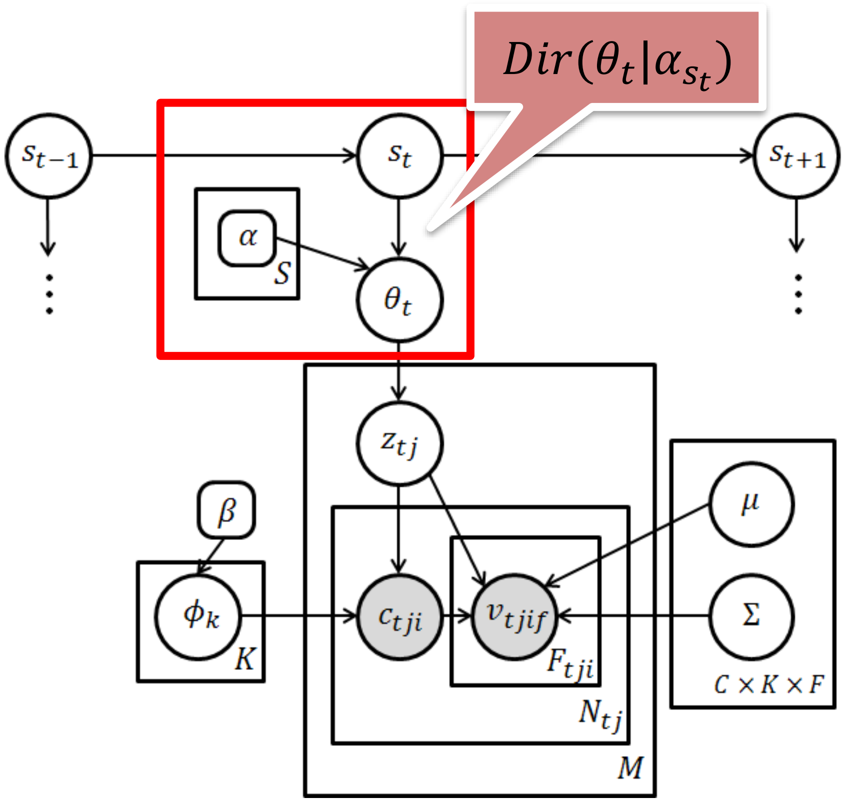
- topic proportion

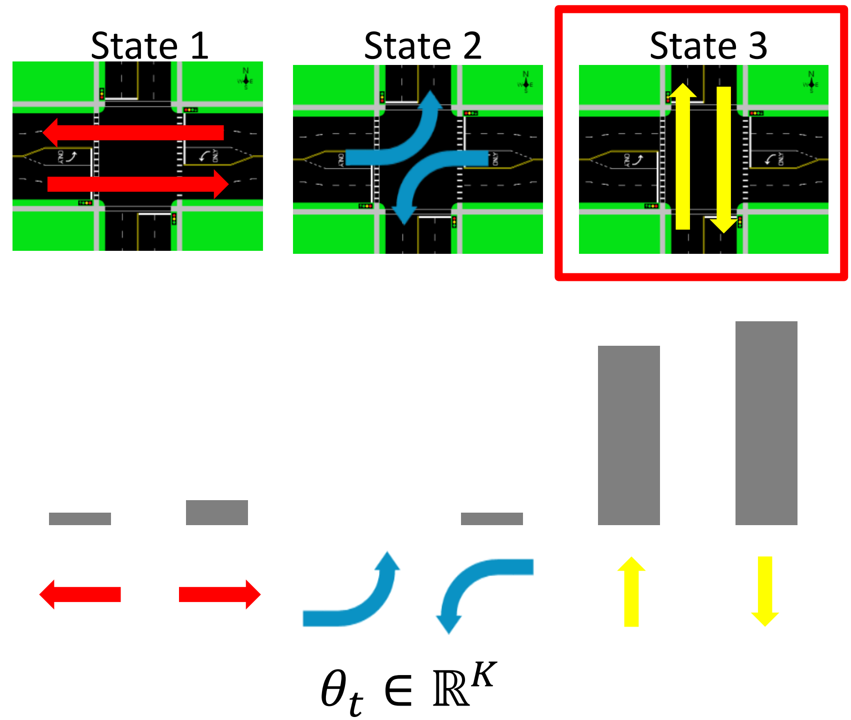




Distribution of Topic Occurrence

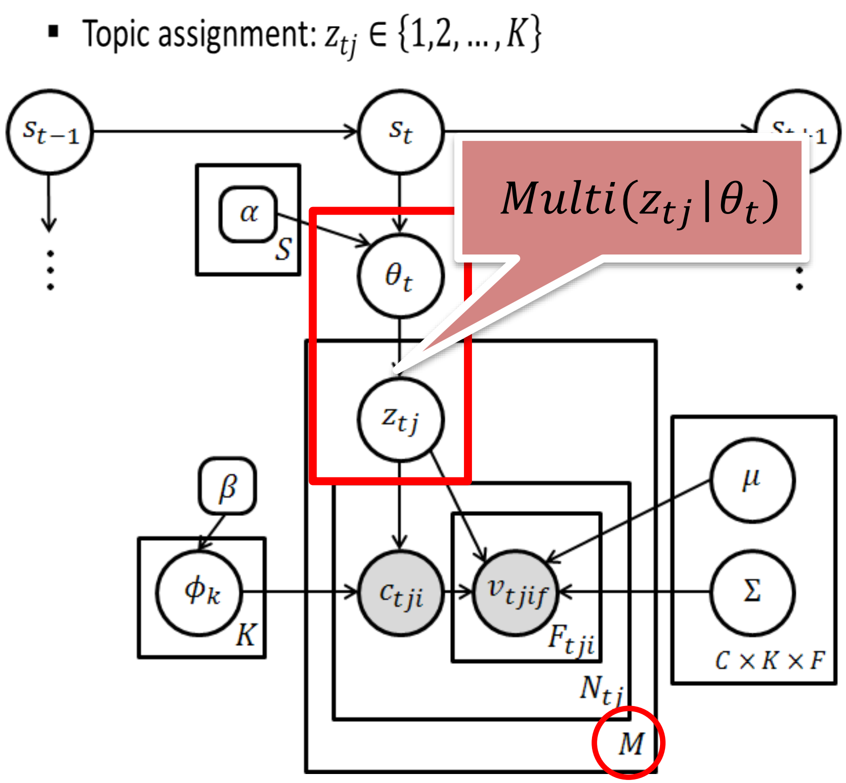
- topic proportion



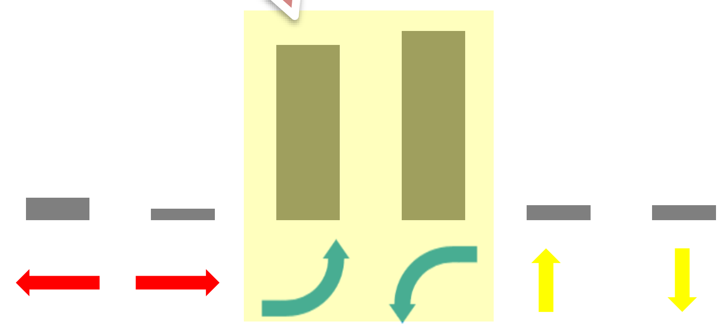


Trajectory patterns

- Topic assignment: $z_{tj} \in \{1,2,\ldots,K\}$



non-conflict patterns



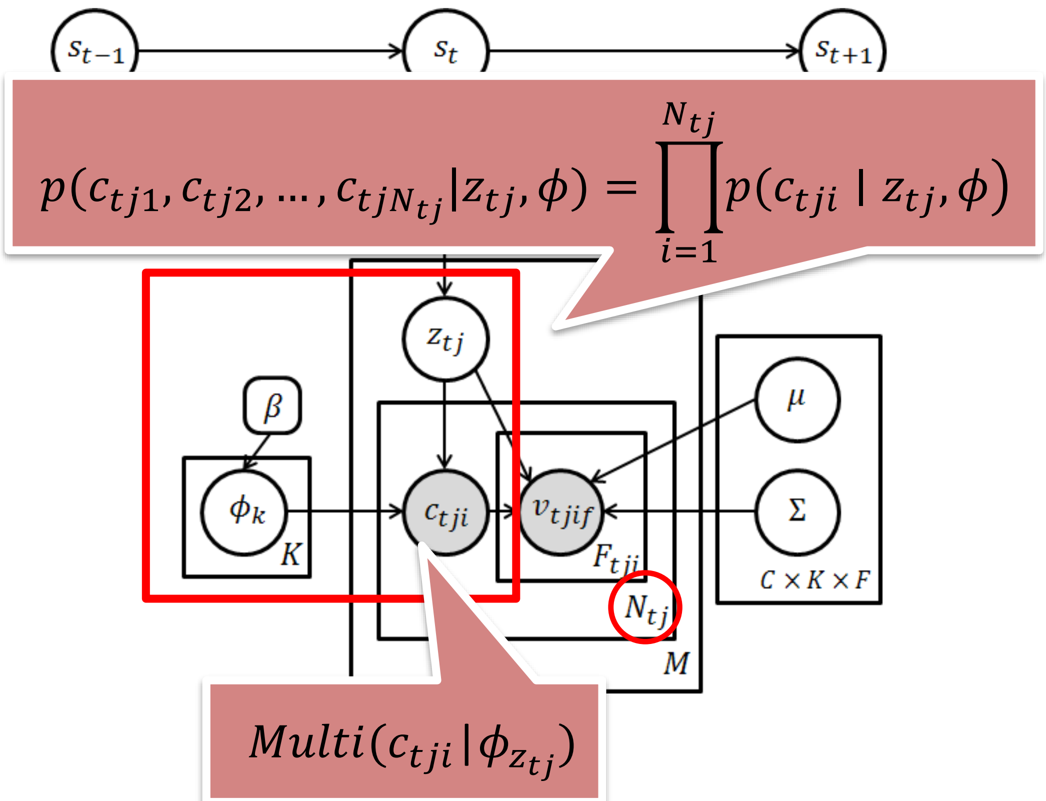
$\theta_t \in \mathbb{R}^K$

sequence of z_{tj} is i.i.d.

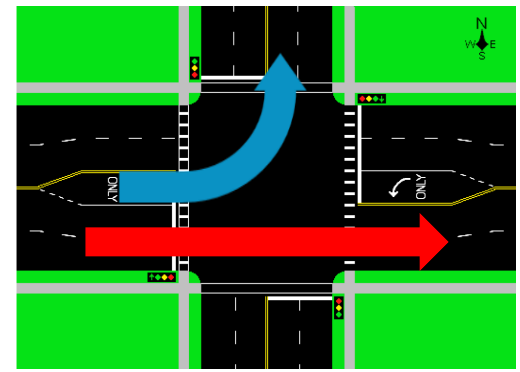
$$p(z_{t1}, z_{t2}, \ldots, z_{tM}|\theta_t) = \prod_{j=1}^{M} p(z_{tj} \mid \theta_t)$$

Association between Cell and Trajectory

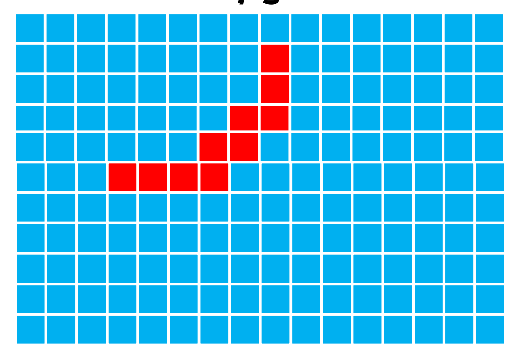
- Neglect temporal dependency among cells



$z_{tj} = 3$

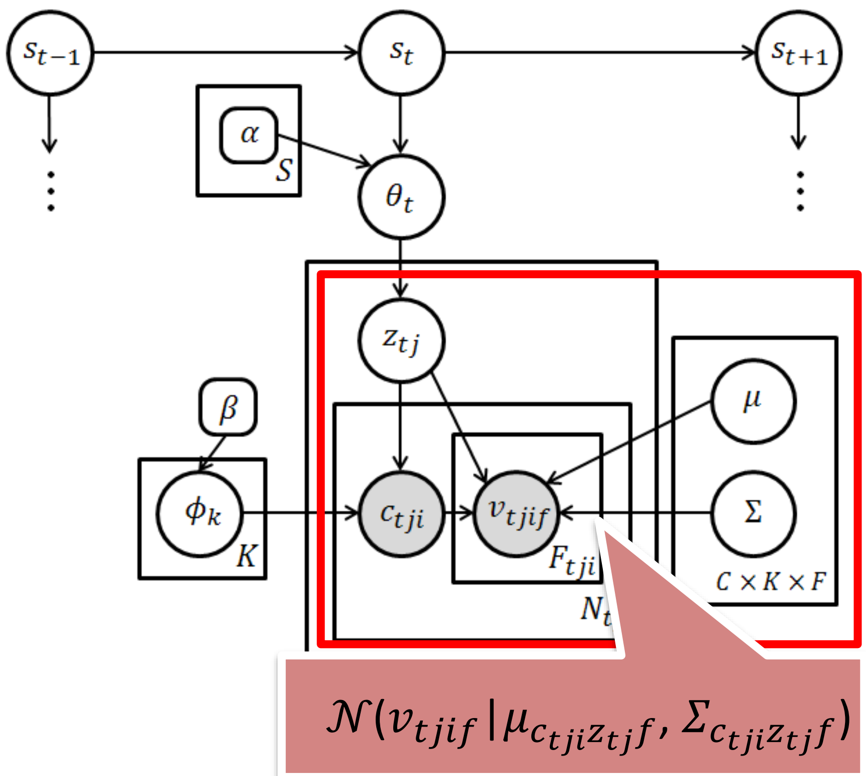


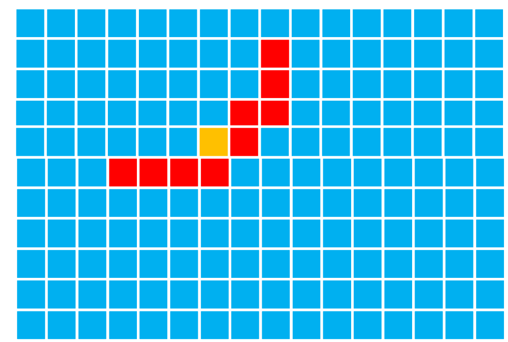
$\boldsymbol{\phi_3} \epsilon \mathbb{R}^C$

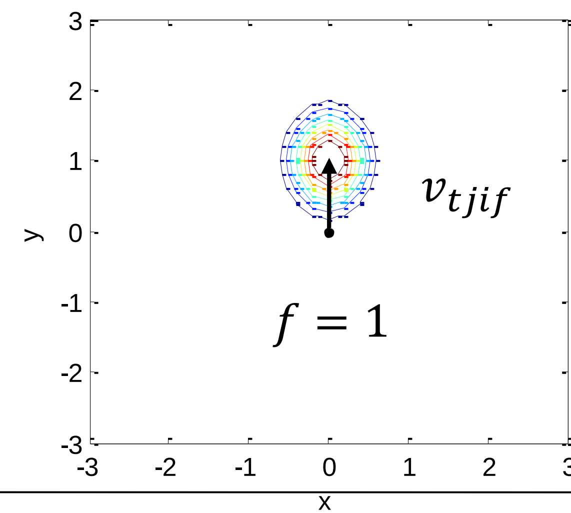


Velocity v_{tjif} Modelling

- Contains temporal information of observations

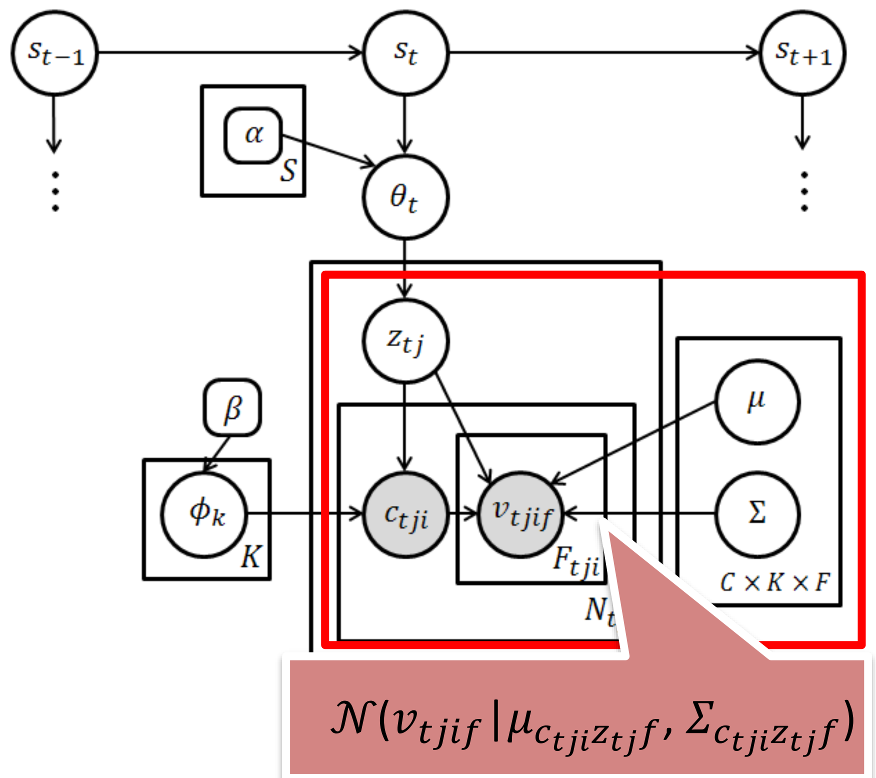


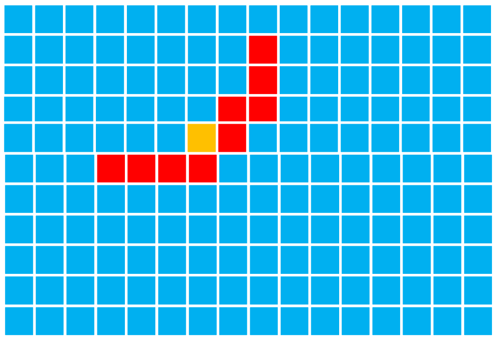


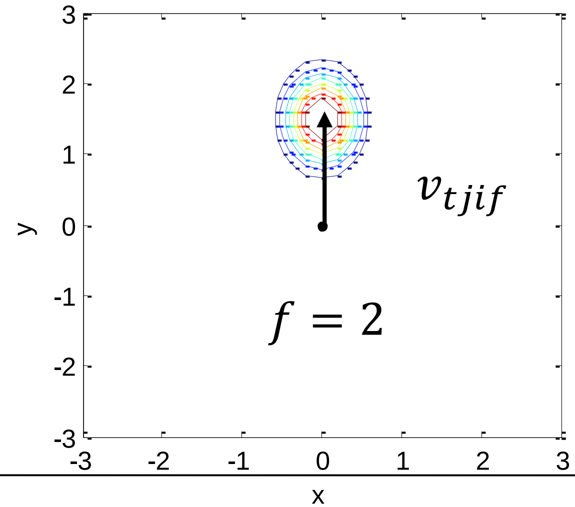


Velocity v_{tjif} Modelling

- Contains temporal information of observations

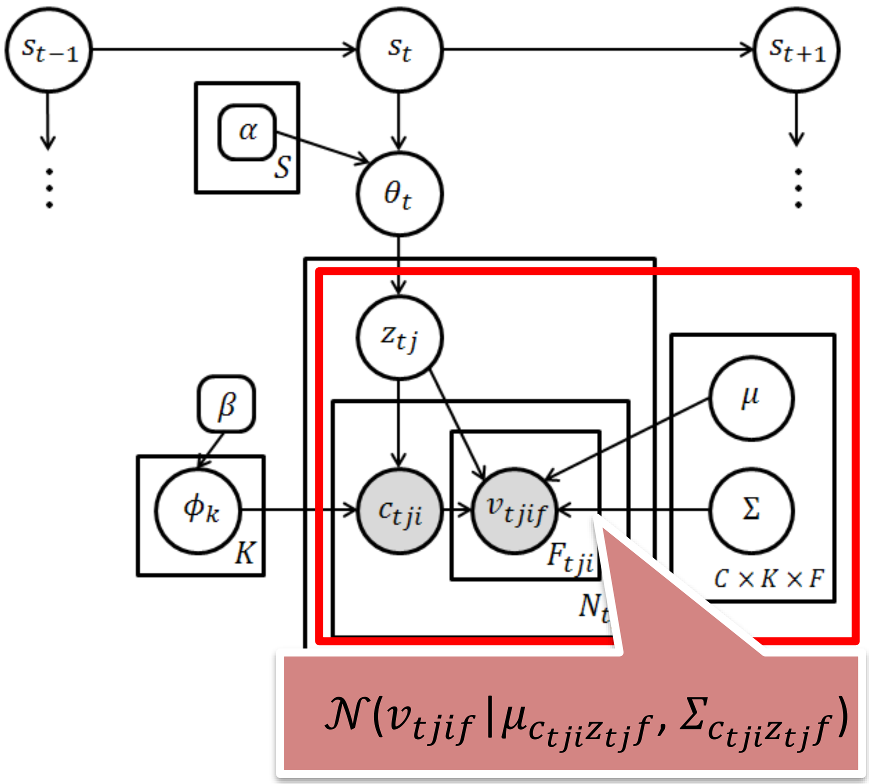


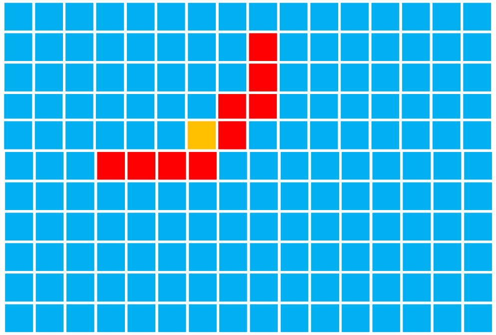


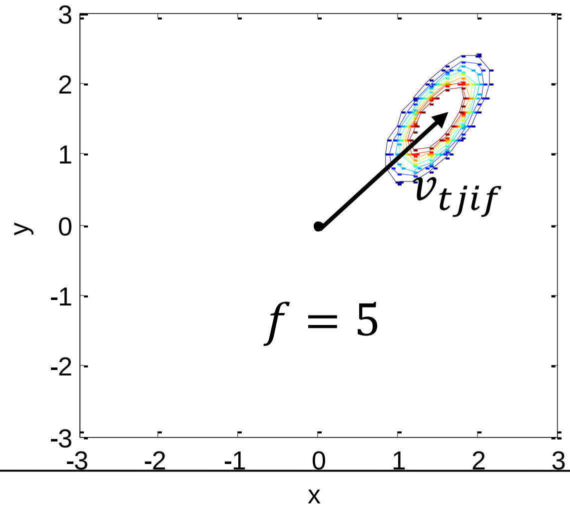


Velocity v_{tjif} Modelling

- Contains temporal information of observations

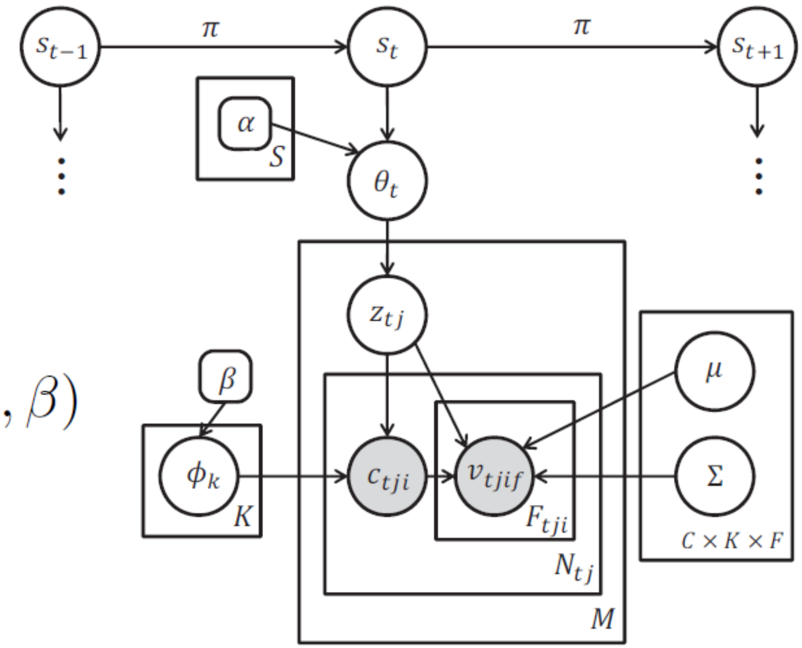






Bayesian Inference (Learning)

- To infer the latent variables and parameters from the given observations $\{c_{tji}\}$, $\{v_{tjif}\}$



$$s^*, \phi^*, \theta^*, z^*, \mu^*, \Sigma^* = \underset{s,\phi,\theta,z,\mu,\Sigma}{\arg\max}\ p(s, \phi, \theta, z, \mu, \Sigma | c, v, \alpha, \beta)$$

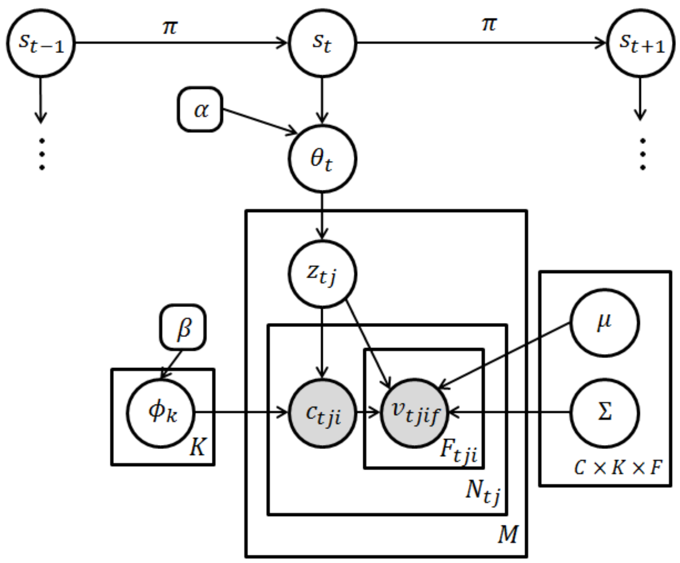
where,

$$p(s, \phi, \theta, z, \mu, \Sigma, c, v|\alpha, \beta) = \left(\prod_{k=1}^{K} p(\phi_k|\beta)\right) \prod_{t=1} p(s_t \mid s_{t-1}) p(\theta_t \mid s_t, \alpha)$$

$$\prod_{j=1}^{M} p(z_{tj} \mid \theta_t) \prod_{i=1}^{N_{tj}} p(c_{tji} \mid z_{tj}, \phi) \prod_{f=1}^{F_{tji}} p(v_{tjif} \mid z_{tj}, c_{tji}, \mu, \Sigma)$$

Online Inference (Learning)

- In case of distributed processing for online learning of topic models, Variational Inference (VI) is better than inference by sampling*
- Applying VI directly to our model is not trivial

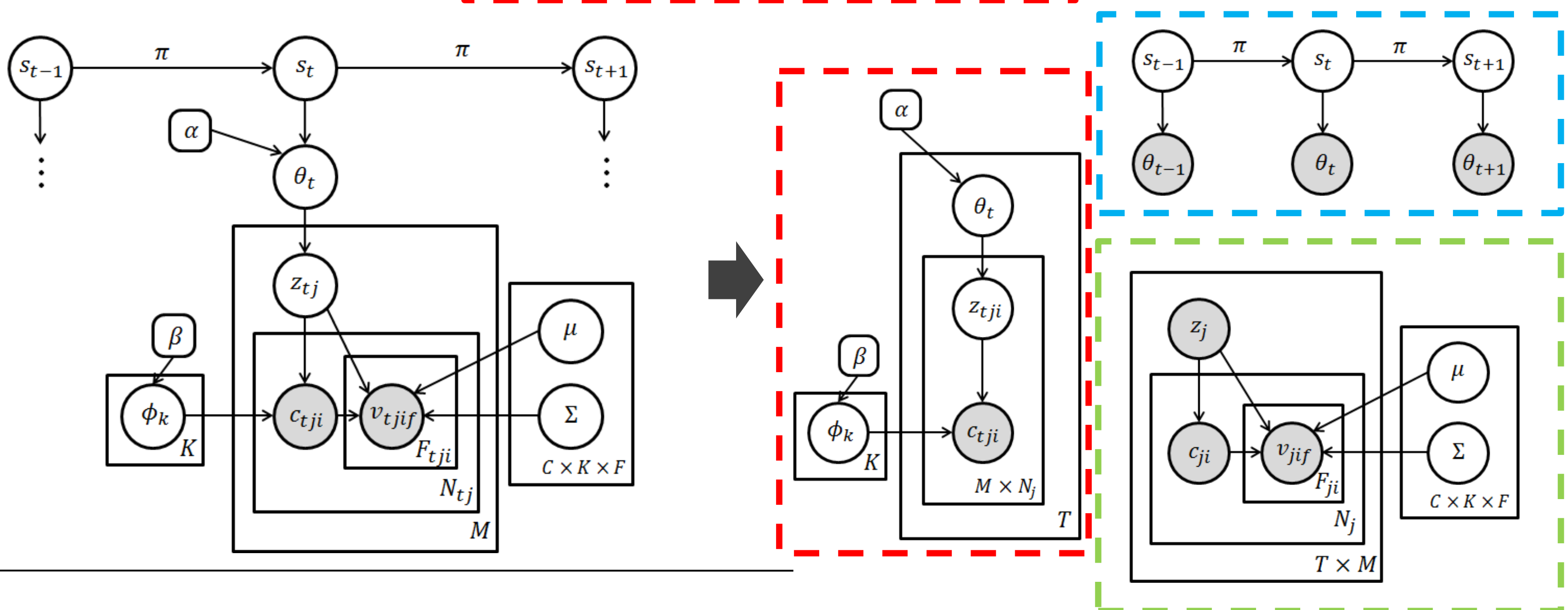


*K. Zhai, J. Boyd-Graber, N. Asadi, and M. Alkhouja. Mr. LDA: A flexible large scale topic modeling package using variational inference in map-reduce. In ACM International Conference on World Wide Web, 2012.

Online Learning

- Divide the model into tractable sub-models

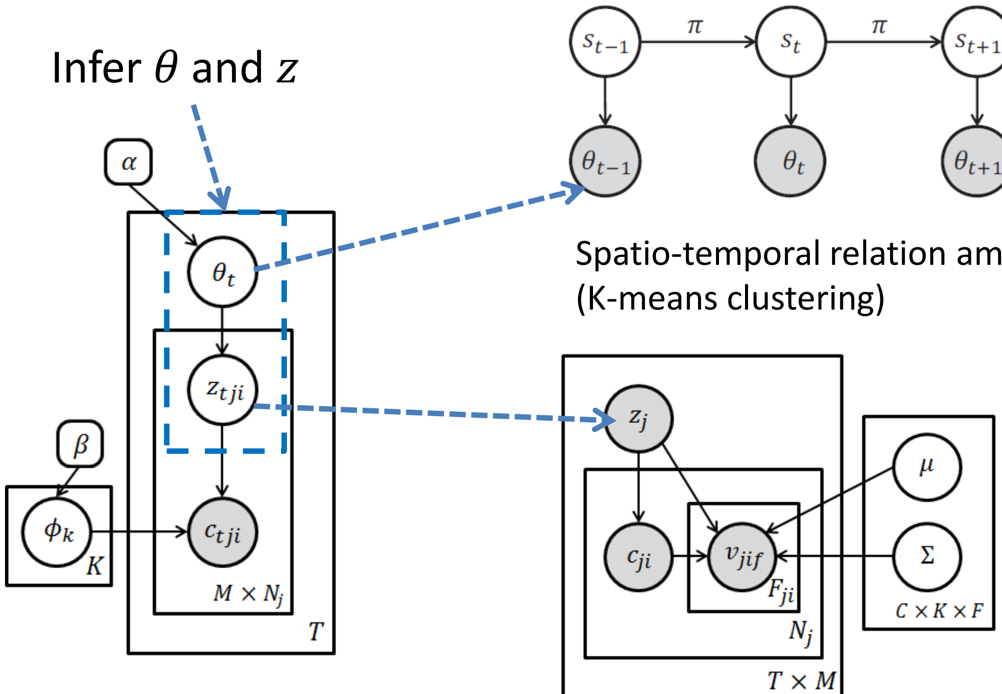
$$p(s,\phi,\theta,z,\mu,\Sigma,c,v|\alpha,\beta) = \left(\prod_{k=1}^{K} p(\phi_k|\beta)\right) \prod_{t=1}^{T} p(s_t \mid s_{t-1})\, p(\theta_t \mid s_t,\alpha) \prod_{j=1}^{M} p(z_{tj} \mid \theta_t) \prod_{i=1}^{N_{tj}} p(c_{tji} \mid z_{tj},\phi) \prod_{f=1}^{F_{tji}} p(v_{tjif} \mid z_{tj}, c_{tji}, \mu, \Sigma)$$



Online Learning

- Divide the model into tractable sub-models

Infer θ and z



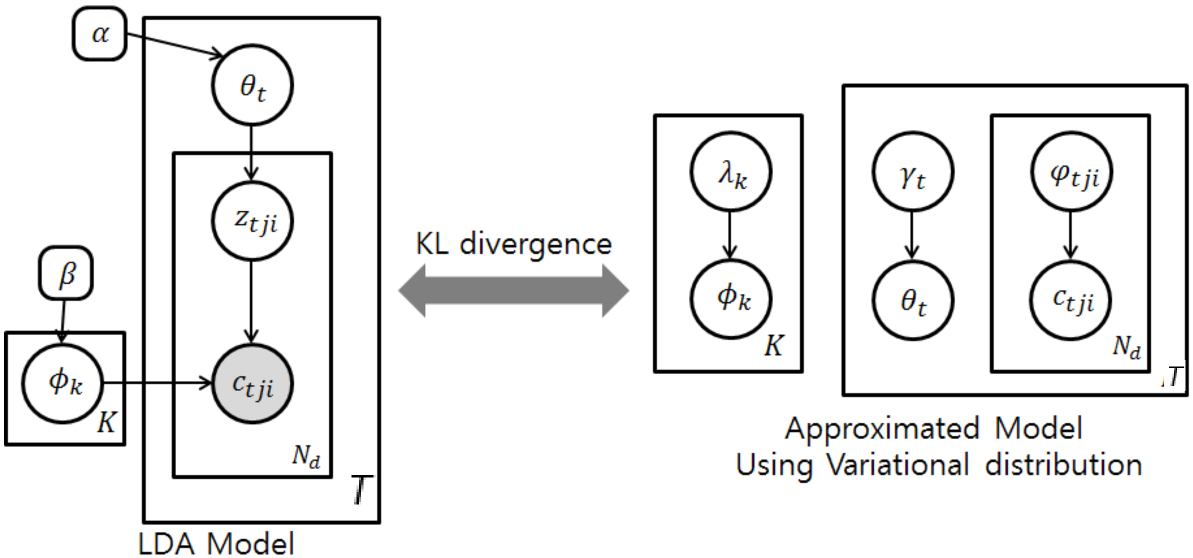
Spatio-temporal relation among activities
(K-means clustering)

Trajectory clustering
(Latent Dirichlet Allocation)

Velocity modeling
(Gaussian models)

Online Trajectory Clustering

- Online variational inference* for LDA (mini-batch)
- The updated parameter is utilized as an initial value in the next mini-batch



$$\widetilde{\lambda}^*, \gamma^*, \varphi^* = \arg\min_{\lambda,\gamma,\varphi} D_{KL}\left(q(\cdot)\,\|p(\cdot)\right)$$

$$where,\quad q(\cdot) = q(\phi, \theta, z|\lambda, \gamma, \varphi),$$

$$p(\cdot) = p(\phi, \theta, z|c, \alpha, \beta).$$

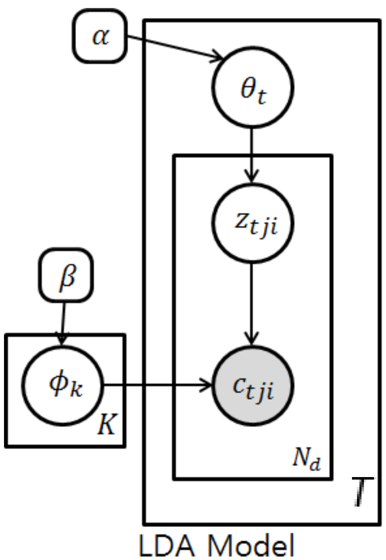
$$q(\{\phi_k\}, \{z_{tji}\}, \{\theta_t\}|\lambda, \varphi, \gamma) =$$

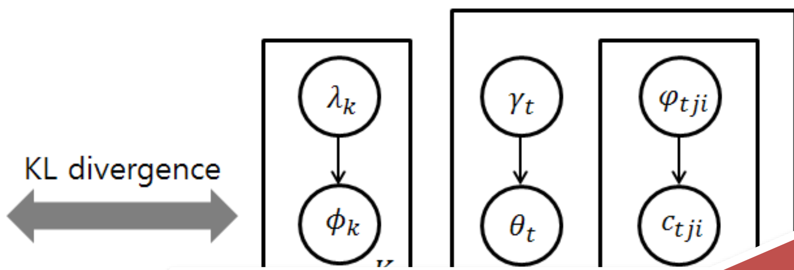
$$\left(\prod_{k=1}^{K} q(\phi_k|\lambda_k)\right)\left(\prod_{t=1}^{T'} q(\theta_t|\gamma_t)\right)\left(\prod_{t,j,i}^{T',M,N_{tj}} q(z_{tji}|\varphi_{tji})\right)$$

*Hoffman, M., Blei, D.M., Bach, F.: Online learning for latent dirichlet allocation. In: NIPS (2010)

Online Trajectory Clustering

- Online variational inference* for LDA (mini-batch)
- The updated parameter is utilized as an initial value in the next mini-batch





$$\widetilde{\lambda}^*, \gamma^*, \varphi^* = \arg\min_{\lambda,\gamma,\varphi} D_{KL}\left(q(\cdot)\,\|p(\cdot)\right)$$

e, $q(\cdot) = q(\phi, \theta, z|\lambda, \gamma, \varphi)$,

$= p(\phi, \theta, z|c, \alpha, \beta)$.

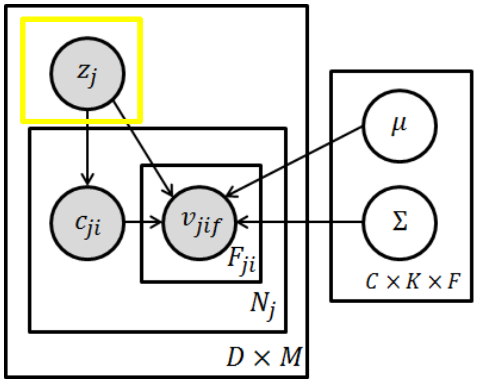
The results are utilized as an initial value in the next mini-batch

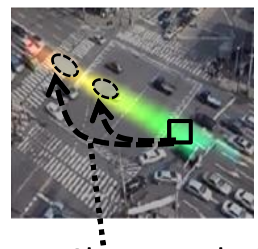
$$q(\{\phi_k\}, \{z_{tji}\}, \{\theta_t\}|\lambda, \varphi, \gamma) = \left(\prod_{k=1}^{K} q(\phi_k|\lambda_k)\right)\left(\prod_{t=1}^{T'} q(\theta_t|\gamma_t)\right)\left(\prod_{t,j,i}^{T',M,N_{tj}} q(z_{tji}|\varphi_{tji})\right)$$

*Hoffman, M., Blei, D.M., Bach, F.: Online learning for latent dirichlet allocation. In: NIPS (2010)

Velocity modeling

- Gaussian parameters are learned for each cell, topic, and time
- Each model can generate previous position of a trajectory





Choose relative positions f-frame ahead using average velocities v_{tjif}

- Online update of Gaussian parameters

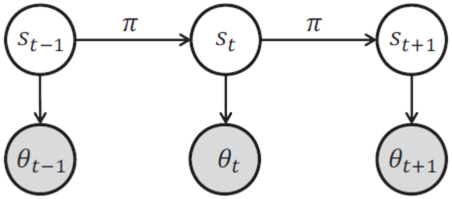
$$\mu_{c_{tji} z^*_{tj} f} = (1 - \rho_{c_{tji} z^*_{tj} f}) \mu_{c_{tji} z^*_{tj} f} + \rho_{c_{tji} z^*_{tj} f} v_{tjif}$$

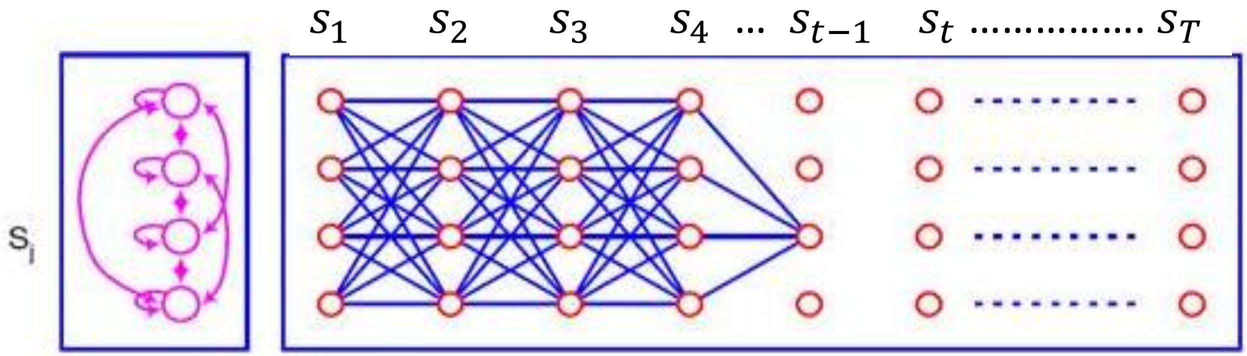
$$Z_{c_{tji} z^*_{tj} f} = (1 - \rho_{c_{tji} z^*_{tj} f}) Z_{c_{tji} z^*_{tj} f} + \rho_{c_{tji} z^*_{tj} f} v_{tjif} v_{tjif}^T$$

$$\Sigma_{c_{tji} z^*_{tj} f} = Z_{c_{tji} z^*_{tj} f} - \mu_{tjif} \mu_{tjif}^T$$

Spatio-temporal relation among activities

- EM approach is used to Infer parameters
- Use θ_t as a vector for clustering
- K-means clustering with K = S

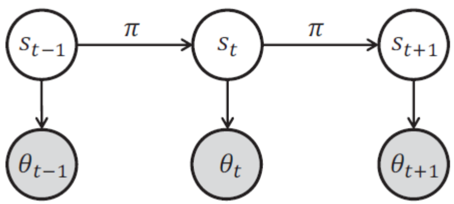




$$\arg\min_{\{\Theta_n\}_{n=1}^{S}} \sum_{n=1}^{S} \sum_{\overline{\theta}^*{}_t \in \Theta_n} \left\| \overline{\theta}^*{}_t - m_n \right\|^2$$

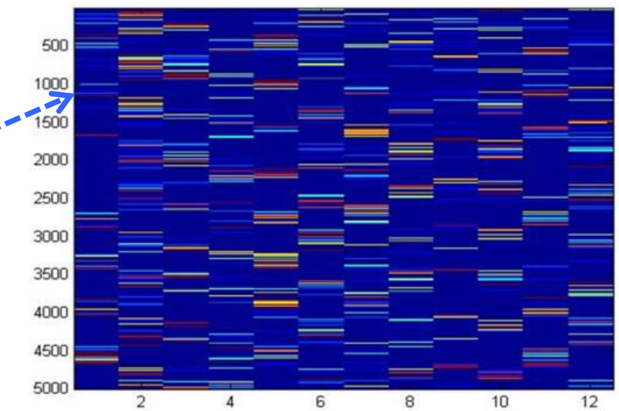
Spatio-temporal relation among activities

- EM approach is used to Infer parameters
- Use θ_t as a vector for clustering
- K-means clustering with K = S



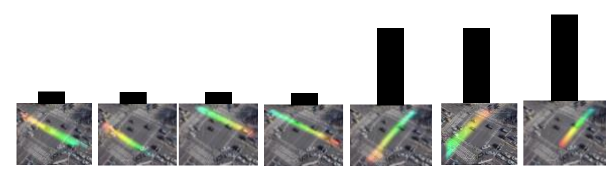
Each row is θ_t

θ_t: K dimensional multinomial



m_1

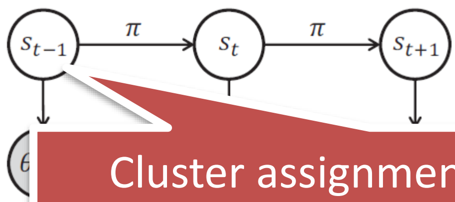
m_2



$$\arg\min_{\{\Theta_n\}_{n=1}^{S}} \sum_{n=1}^{S} \sum_{\overline{\theta}^*{}_t \in \Theta_n} \left\| \overline{\theta}^*{}_t - m_n \right\|^2$$

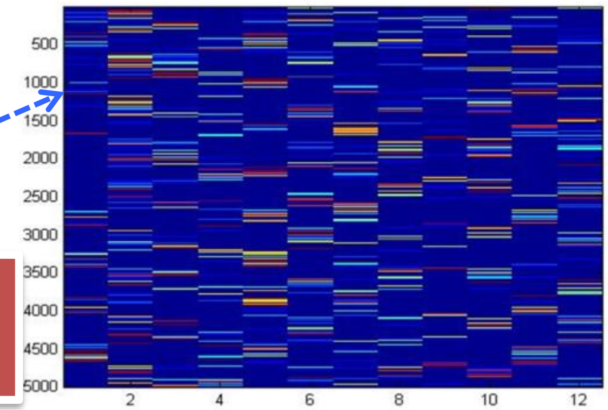
Spatio-temporal relation among activities

- EM approach is used to Infer parameters
- Use θ_t as a vector for clustering
- K-means clustering with K = S

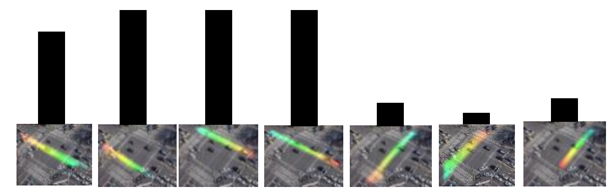


Cluster assignment corresponds to the state

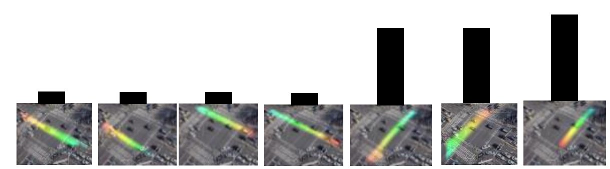
θ_t: K dimensional multinomial



m_1



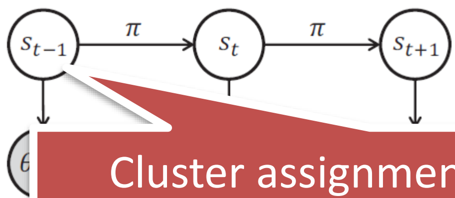
m_2



$$\arg\min_{\{\Theta_n\}_{n=1}^{S}} \sum_{n=1}^{S} \sum_{\overline{\theta}^*{}_t \in \Theta_n} \left\| \overline{\theta}^*{}_t - m_n \right\|^2$$

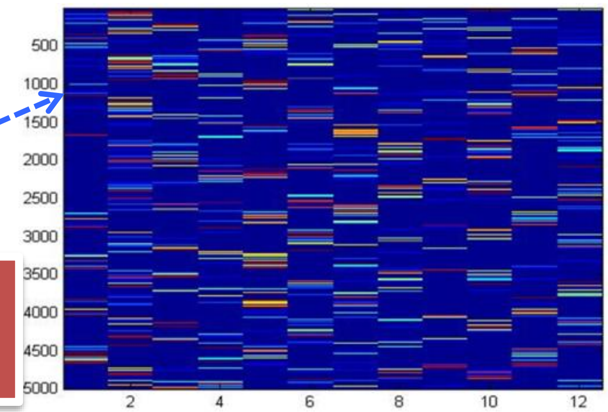
Spatio-temporal relation among activities

- EM approach is used to Infer parameters
- Use θ_t as a vector for clustering
- K-means clustering with K = S



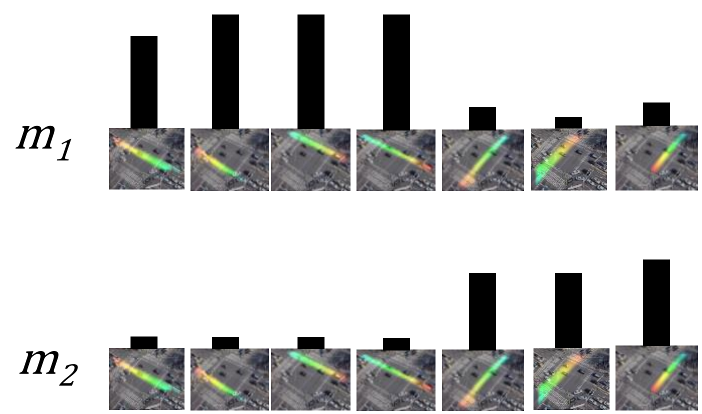
Cluster assignment corresponds to the state

θ_t: K dimensional multinomial



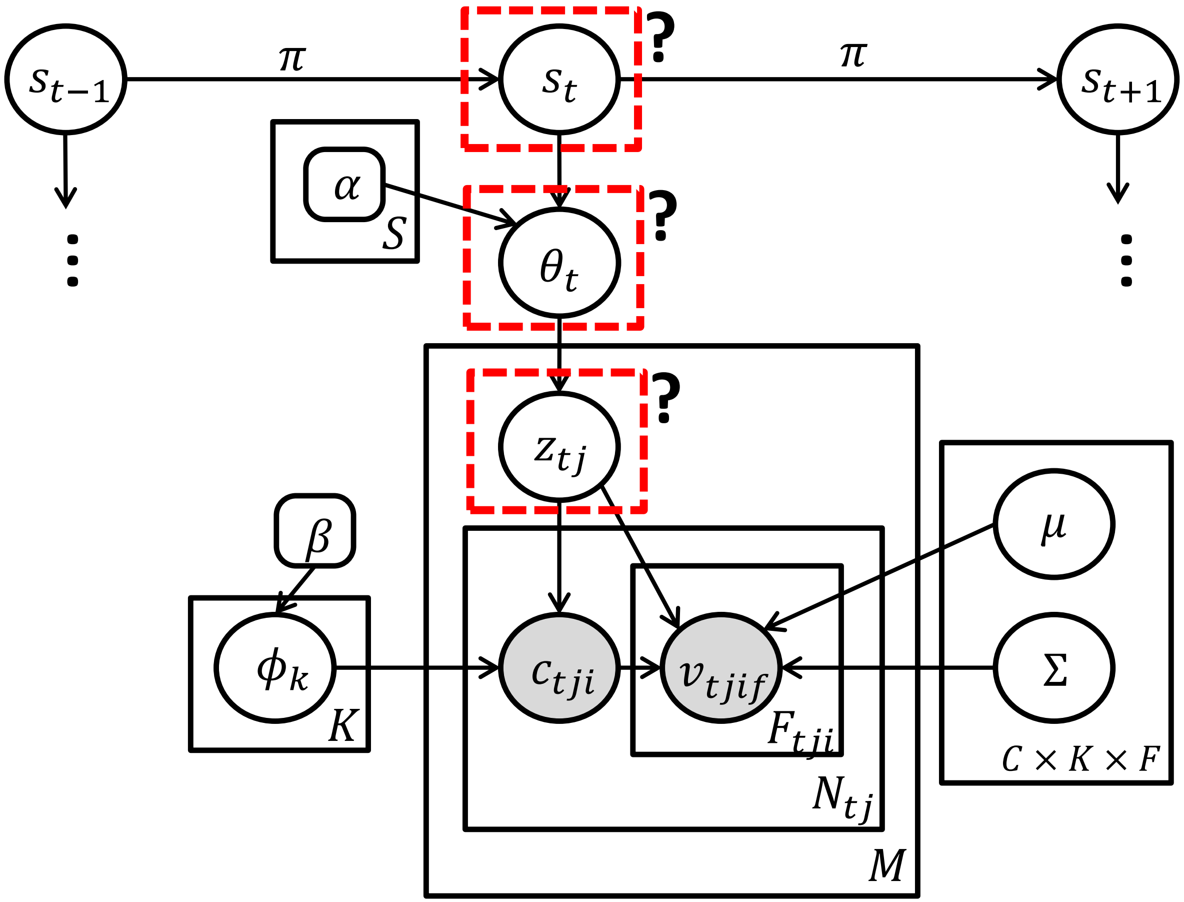
m_n implies representative patterns about spatial co-occurrences of trajectory patterns

$$\arg\min_{\{\Theta_n\}_{n=1}^{S}} \sum_{n=1}^{S} \sum_{\overline{\theta}^*{}_t \in \Theta_n} \left\| \overline{\theta}^*{}_t - m_n \right\|^2$$



Anomaly Test

- Use the distribution parameters inferred from the learning phase



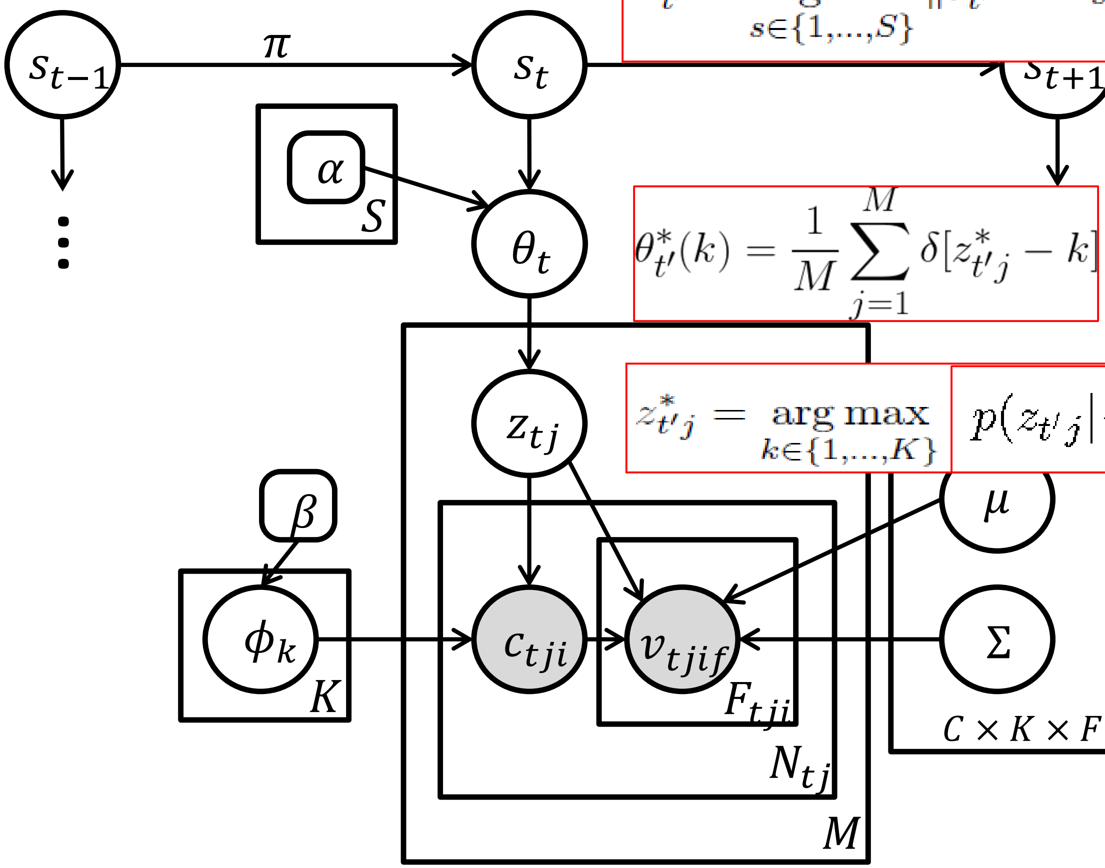
Anomaly Test

- For the approximation, the same assumptions are used as in case of learning phase

$$s^*_{t'} = \arg\min_{s\in\{1,\ldots,S\}} \|\theta^*_{t'} - m_s\|$$

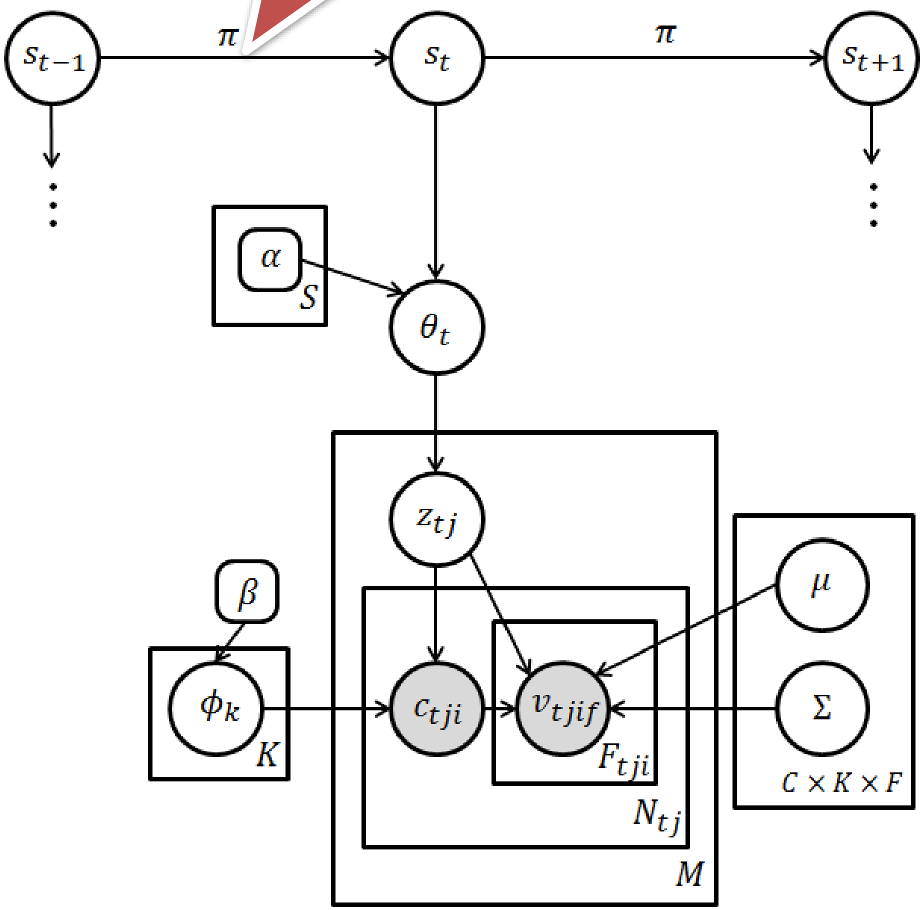
$$\theta^*_{t'}(k) = \frac{1}{M}\sum_{j=1}^{M} \delta[z^*_{t'j} - k]$$

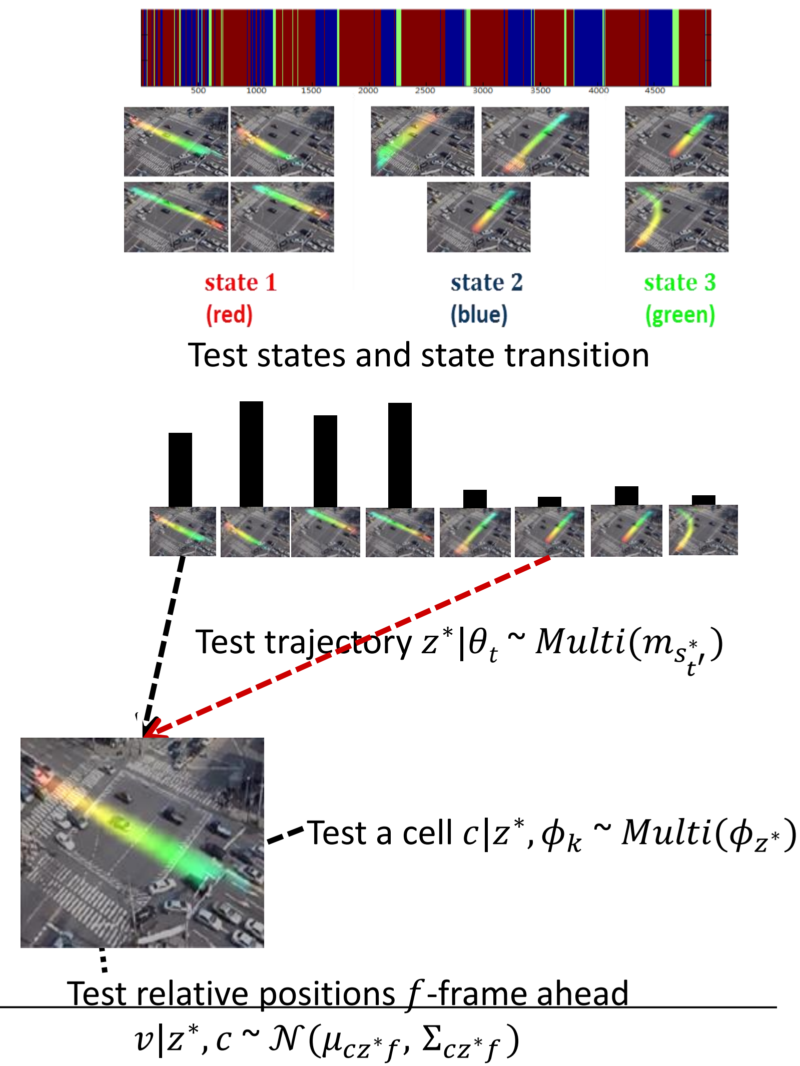
$$z^*_{t'j} = \arg\max_{k\in\{1,\ldots,K\}} p(z_{t'j} | \{c_{t'ji}\}_{i=1}^{N_{t'j}}, \phi)$$



An

Examine the temporal relation among the typical patterns of trajectories

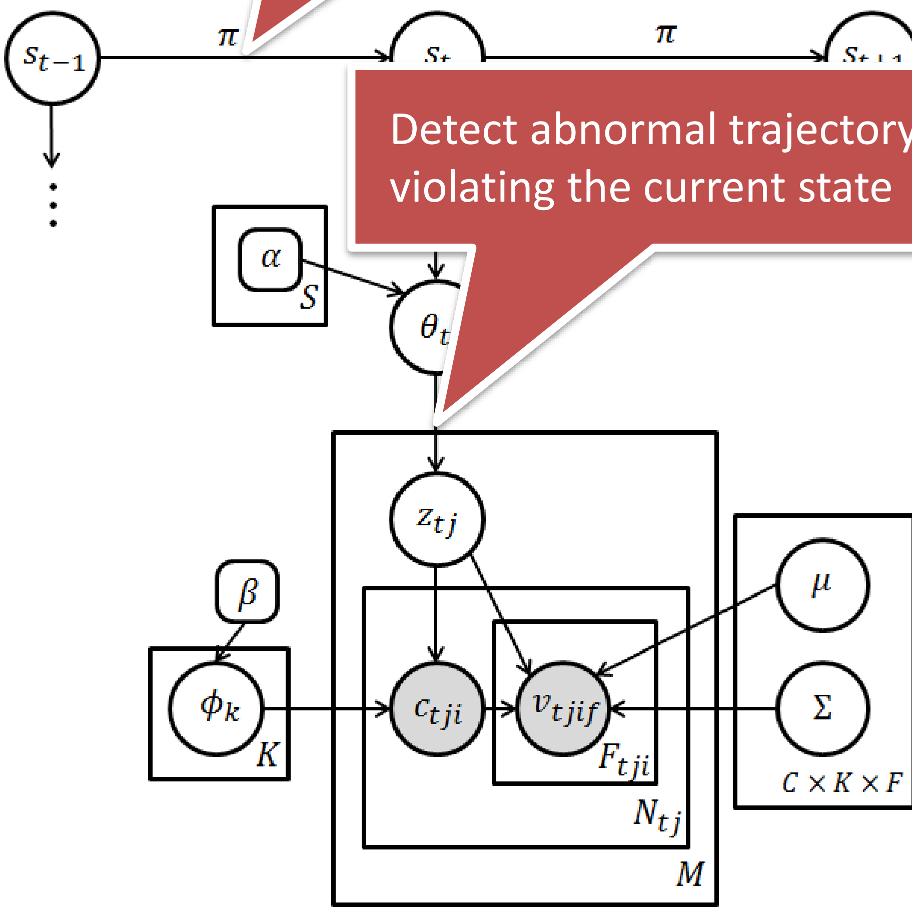


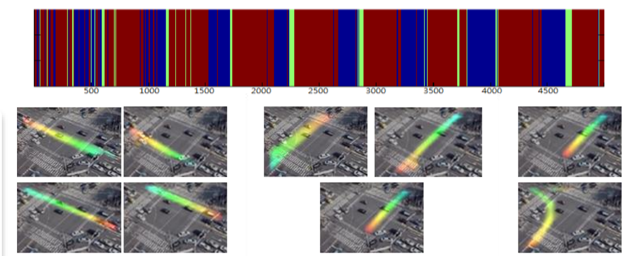


An

Examine the temporal relation among the typical patterns of trajectories

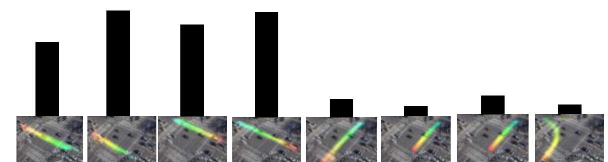
Detect abnormal trajectory violating the current state





state 1 (red) | state 2 (blue) | state 3 (green)

Test states and state transition



Test trajectory $z^*|\theta_t \sim Multi(m_{s^*_{t'}})$

Test a cell $c|z^*, \phi_k \sim Multi(\phi_{z^*})$

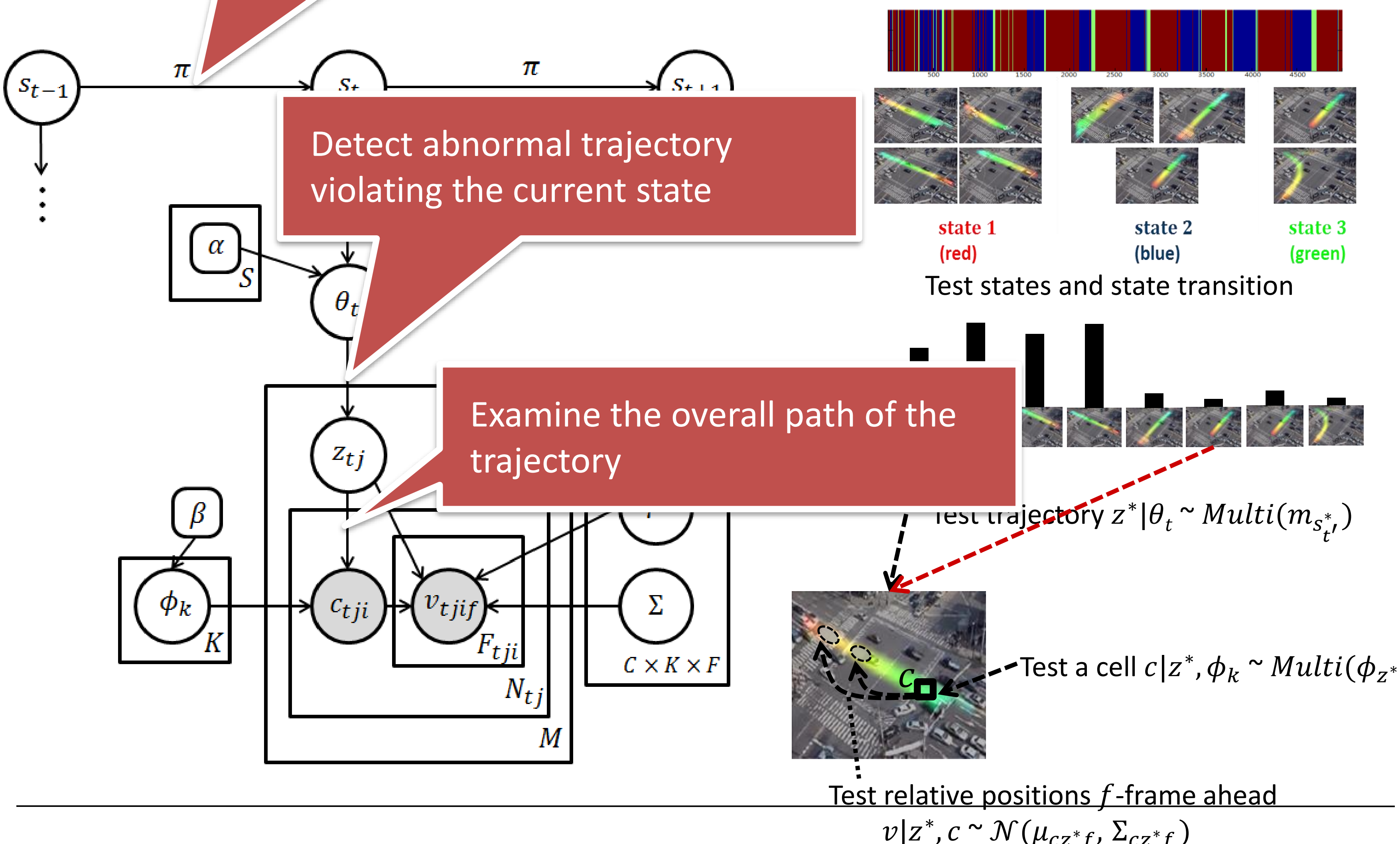
Test relative positions f-frame ahead
$v|z^*, c \sim \mathcal{N}(\mu_{cz^*f}, \Sigma_{cz^*f})$

An

Examine the temporal relation among the typical patterns of trajectories

Detect abnormal trajectory violating the current state

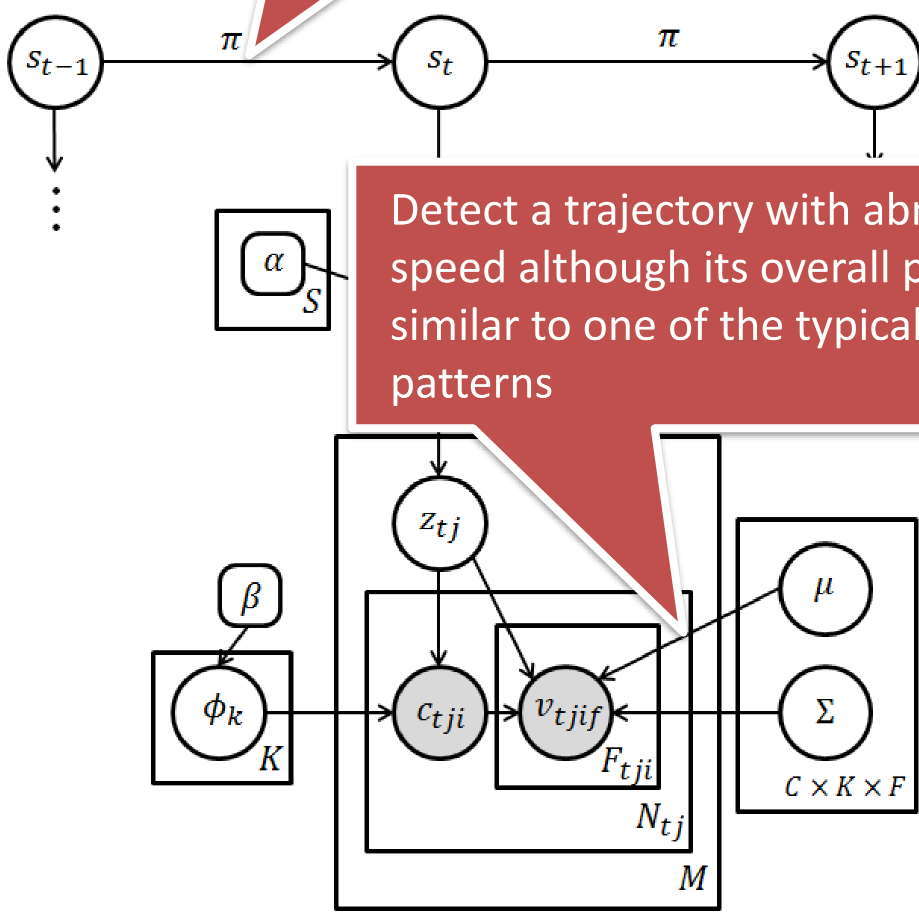
Examine the overall path of the trajectory

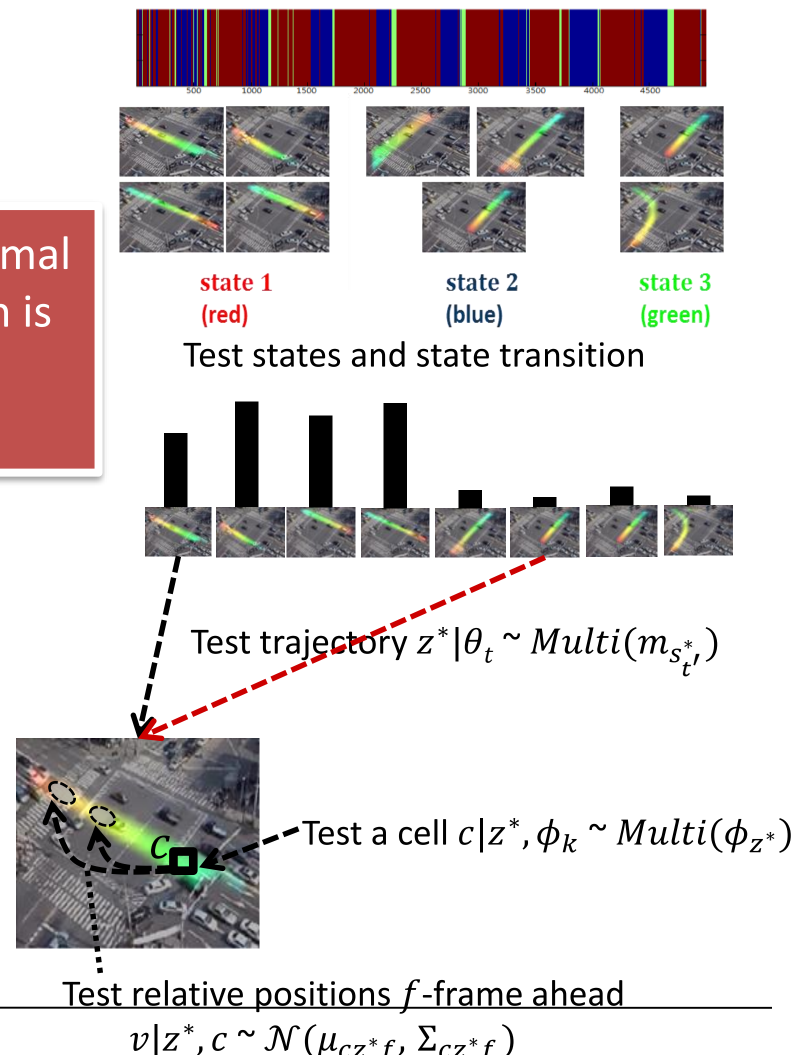


An

Examine the temporal relation among the typical patterns of trajectories

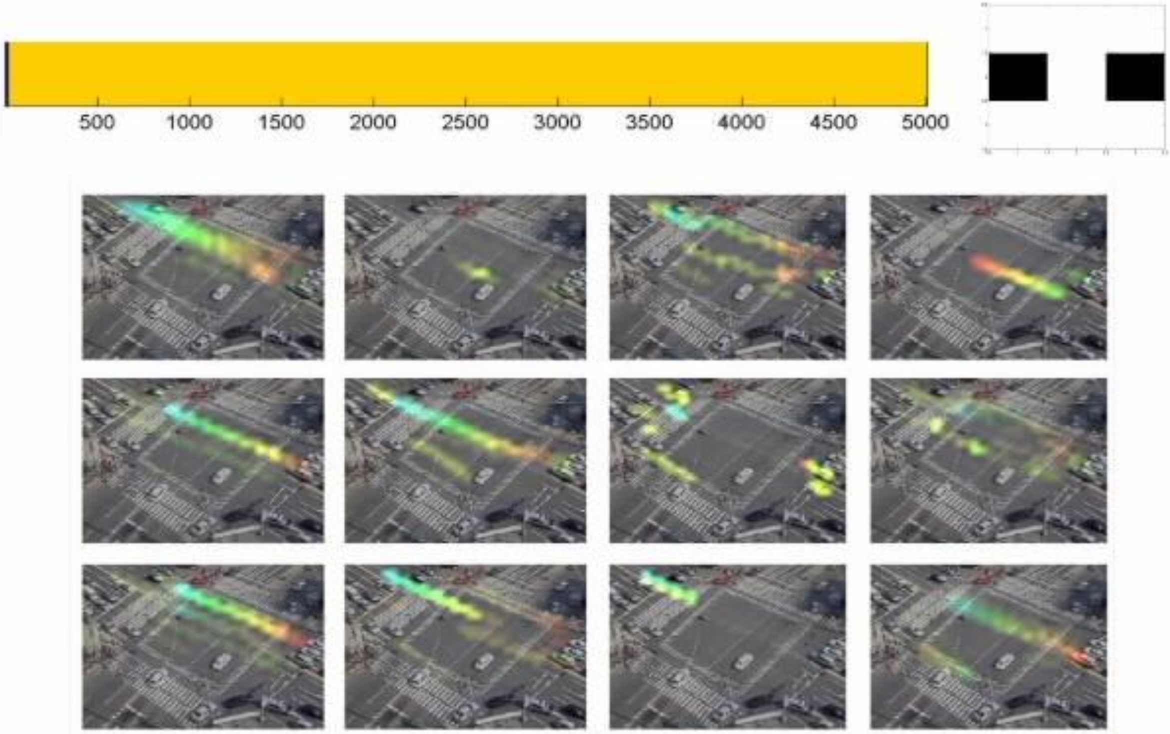
Detect a trajectory with abnormal speed although its overall path is similar to one of the typical patterns





Experimental Results

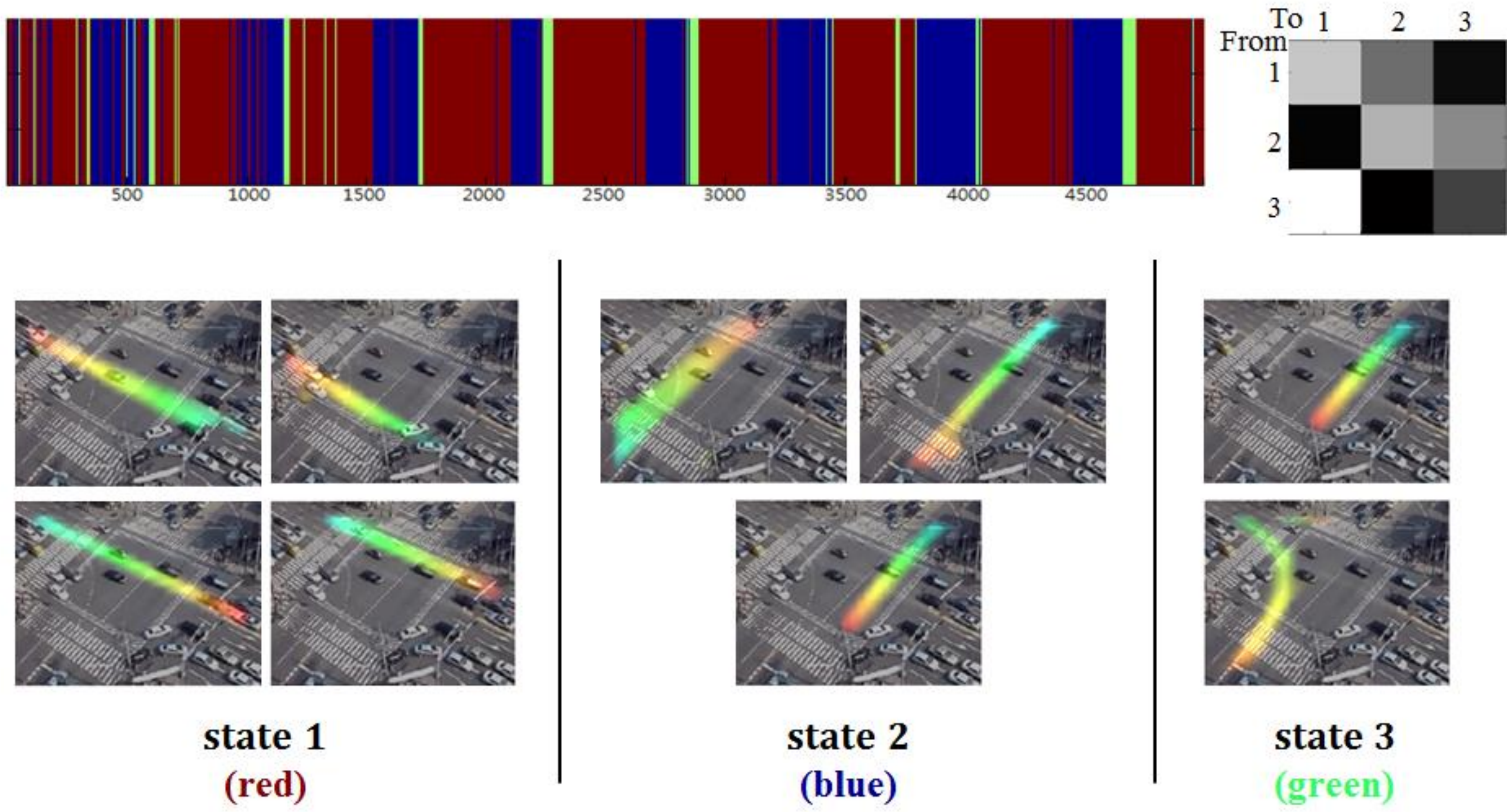
- Online learning sequence



Online inference

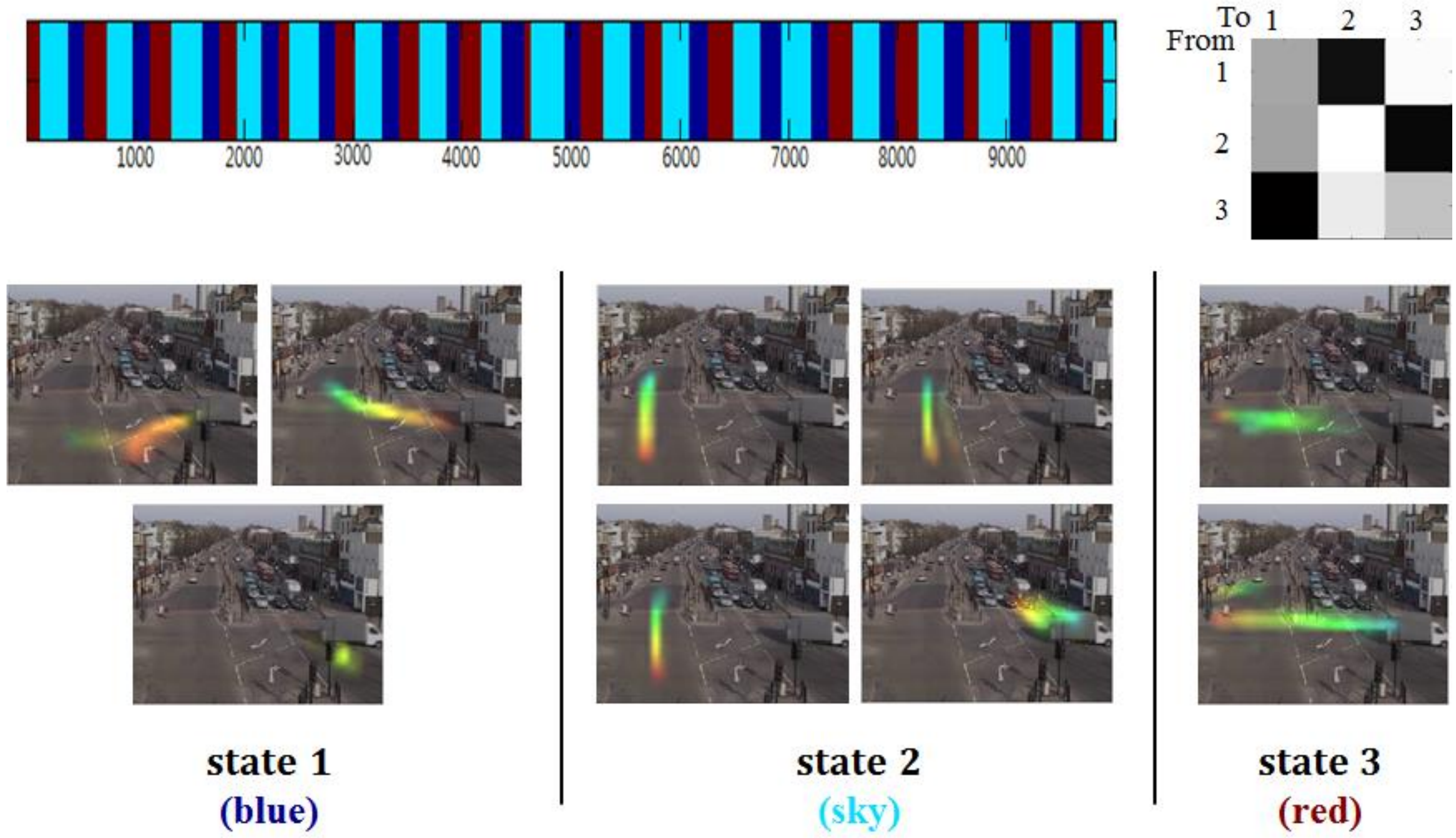
Experimental Results

- Motion pattern model (qualitative result)



Experimental Results

- Motion pattern model (qualitative result)



Experimental Results

- Anomaly detection (qualitative result)

Interim Summary

- What is variational inference ?
- Kullback–Leibler divergence (KL-divergence) formulation
- Dual of KL-divergence
- Variational Inference for LDA
- Estimating variational parameters
- Estimating LDA parameters
- Application of VI to Generative Image Modeling
- Application of LDA toTraffic Pattern Analysis

Summary of Course

Bayes Rule | Likelihood | Posteriori | Priori | Bayes Decision

Learning	Generative Model	Discriminative Model	Learning
ML(P)E Bayes. L	GM	Lin. Classifier SVM	LS Convex O.
Histogram	K-NN Parzen W.	Random Forest	Entropy
EM, MCMC	GMM	K-SVM K-SVDD	Convex O.
MCMC, VI	Bayesian Net	Deep NN	SA GA BP(GD) NM
MLE VI	Boltzm. Machine	ICA	K-L Divergence
MCMC, VI	Latent DA	Linear DA	Max. Separa.
EM	HMM	PCA	Max. Scatter