Supplementary material (inference and sampling equations) for "*Bridging the Past, Present and Future; Modeling Scene Activities From Event Relationships and Global Rules*"

Jagannadan Varadarajan[1,2], Rémi Emonet[1] and Jean-Marc Odobez[1,2]
[1] Idiap Research Institute, CH-1920, Martigny, Switzerland
[2] École Polytechnique Fédéral de Lausanne, CH-1015, Lausanne, Switzerland
{vjagann,remonet,odobez}@idiap.ch

This material has two sections. The first section details the sampling equations used for inference in the model presented in the submitted paper. In the second section we provide more experimental results that could not be included in the main paper due to space restrictions.

1. Generative model and notations

The graphical model is reproduced in Fig. 1. To make this a self-contained, we first recall the generative model, its notations and then provide the inference derivation. It corresponds to the following set of equations:

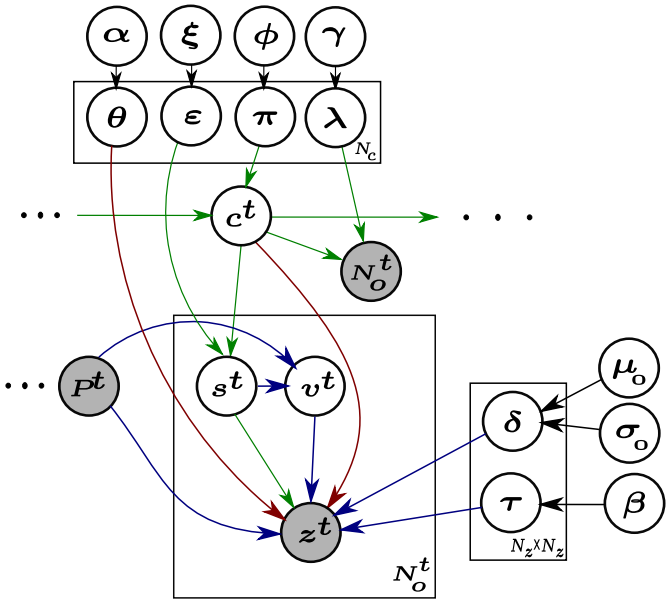


Figure 1: Mixed Event Relationship Model: Shaded cicles represent observed variables. The arrows represent dependencies between variables. Colors green - blue - red are used to indicate dependencies used for a)all events, b)dependent and c)independent events respectively.

2. Notations

3. Generative Process

The method of generating the event occurrence matrix is described as follows.

Symbol	Description
α	Dirichlet prior on independent events
β	Dirichlet prior on motif transitions
μ_0, σ_0^2	Hyper-parameters of Gaussian prior
$\gamma = \{\gamma_1, \gamma_2\}$	Hyper-parameters of Gamma prior
$\xi = \{\xi_0, \xi_1\}$	Hyper-parameters of Beta prior
φ	Dirichlet prior on state transitions
$\tau_{z'}(z)$	Transition between event z' to z
$\delta_{z',z}, \sigma^2$	Mean and a fixed variance on time lag between z' and z
θ_k	Distribution on motifs, for each state k
λ_k	Poisson parameter to select N_o^t, for each state k
ϵ_k	Bernoulli parameter to select s_i^t, for each state k
π_k	Global state transitions, for each state k
N_o^t	Number of events at time t
N_o	Total number of events at all times
P^t	Set of past events for time t
N_p^t	Number of past events for time t
N_z	Number of events
N_c	Number of global scene states
T_k	Number of time instants explained by state k
T_d	Number of time steps in the event occurrence matrix
T_l	Maximum lag allowed for activity relations

Table 1: Notations used in the paper

- for each $k = 1, \ldots, N_c$ global states;
  - draw Poisson parameter $\lambda_k \sim \text{Gamma}(\gamma_1, \gamma_2)$
  - draw Bernoulli parameter $\epsilon_k \sim \text{Beta}(\xi_0, \xi_1)$
  - draw Multinomial parameter $\theta_k \sim \text{Dirichlet}(\alpha)$
  - draw state transitions $\pi_k \sim \text{Dirichlet}(\varphi)$
- for each event type $z \in [1..N_z]$, draw transitions $\tau_z \sim \text{Dirichlet}(\beta)$;
- for each event type pair $z', z \in [1..N_z]^2$, draw lags $\delta_{z',z} \sim \mathcal{N}(\delta_{z',z}|\mu_0, \sigma_0^2)$
- for each t in $1, \ldots, T_d$
  - draw $c^t \sim \text{Discrete}(\pi_{c^{t-1}})$
  - draw a number $N_o^t \sim \text{Poisson}(\lambda_{c^t})$
  - for each i in $1, \ldots, N_o^t$
  - draw a binary value $s_i^t \sim \text{Bernoulli}(\epsilon_{c^t})$
  - draw $O_i^t \sim \text{Discrete}(\theta_{c^t})$ if $s_i^t = 1$,
  - draw $v_i^t | P^t \sim \text{Uniform}(\frac{1}{N_p^t})$, if $s_i^t = 0$ and, draw $O_i^t \sim \mathcal{N}(t - t'|\delta_{z',z}, \sigma) \cdot \text{Discrete}(\tau_{z'})$, where $P^t(v_i^t)$ occurs at time $t' < t$ and is of type z'. σ is a fixed variance for all pairs of events.

An important step in the generative process is the sampling of the decision variable s_i^t. In cases where $s_i^t = 1$, the indicator variable v_i^t assumes a value indicating a past event in the set P^t. In cases where $s_i^t = 1$, v_i^t is given a dummy value of -1. i.e,

$$v_i^t = \begin{cases} j \in [1, \ldots, N_p^t] & \text{if } s_i^t = 0, \text{ (dependent case)} \\ -1 & \text{otherwise, (independent case)} \end{cases} \quad (1)$$

4. Gibbs Sampling

As is the case of many hierarchical Bayesian models like [2, 3], exact inference for our model is intractable. But since the model consists of conjugate pairs like Gamma-Poisson, Dirichlet-Multinomial and Beta-Bernoulli and Normal-Normal, it is possible to derive a collapsed Gibbs sampling algorithm by integrating out the parameters $\{\pi, \lambda, \epsilon, \theta, \tau, \delta\}$. The algorithm proceeds by iteratively sampling the decision variable s_i^t, indicator variable v_i^t for each observation O_i^t conditioned on all other variables, parameters and hyper-parameters. The state indicators c^t are sampled for each time instant conditioning on rest of the variables.

Since each occurrence also gives its occurrence time implicitly, we will drop the t associated with s_i^t, v_i^t, O_i^t and simplify them by using s_i, v_i and o_i instead except in places where time needs to be mentioned explicitly. We will also use capital letters O, S, V to refer to the set of occurrences, their corresponding selector variables, and indicator variables. O_{-i}, S_{-i}, V_{-i}, will indicate all the occurrences, selector variables and the indicator variables except the i^{th} one. C^{-t} is used to indicate the state variables at all time instants except at the current time i.e., c^t. The set of hyper-parameters set $\{\varphi, \gamma, \xi, \alpha, \beta, \mu_0, \sigma_0^2\}$ is simply referred as hp.

4.1. Selector and Indicator variables

We would like to sample the selector variables s_i and v_i together for each observation o_i. We need to sample this according to the probabilities of four different conditions.

- Case 1: $p(s_i = 1, v_i = -1)$: When $s_i = 1$, v_i takes a dummy value of -1. This probability depends on two factors: i) Probability of seeing a $s_i = 1$ at the current state and ii) Probability of the event corresponding to the current observation to be generated by the current state directly.

$$
\begin{aligned}
p(s_i = 1, v_i = -1 | S_{-i}, V, O, C, hp..) &\propto p(s_i = 1, v_i = -1, S_{-i}, V_{-i}, O, C | hp) \\
&\propto p(v_i = -1 | s_i = 1) p(o_i = z | s_i = 1, c^t = k, S_{-i}, V_{-i}, O_{-i}, hp) \cdot \\
&\quad p(s_i = 1 | S_{-i}, c^t = k, C^{-t}, hp) \qquad (2) \\
&\propto p(o_i = z | c^t = k, s_i = 1, S_{-i}, V_{-i}, O_{-i}, hp) \cdot \\
&\quad p(s_i = 1 | c^t = k, S_{-i}, C^{-t}, hp) \qquad (3) \\
&\propto \int p(o_i = z | s_i = 1, c^t = k, \theta_k) p(\theta_k | S_{-i}, O_{-i}, C^{-t}) \, \mathrm{d}\theta_k \cdot \\
&\quad \int p(s_i = 1 | c^t = k, \epsilon_k) p(\epsilon_k | S_{-i}, C^{-t}) \, \mathrm{d}\epsilon_k \qquad (4)
\end{aligned}
$$

  Expression 2 is obtained by splitting the LHS of the equation into $p(o_i = z | s_i = 1, c^t = k, S_{-i}, V_{-i}, O_{-i}, hp)$ which is the likelihood part and $p(s_i = 1 | S_{-i}, c^t = k, C^{-t}, hp)$ which is the prior part and by noting that $p(v_i = -1 | s_i = 1) = 1$.

  Finally the two parts of expression 4 are Multinomial-Dirichlet, and Bernoulli-Beta pairs respectively. They are conjugate pairs and therefore the parameters can be integrated out in closed form. Evaluating the integral leads to calculating the expectation of the Dirichlet and Beta distributions which is given as:

$$
p(s_i = 1, v_i = -1 | S_{-i}, V, O, C, hp..) \propto \frac{q_{-i,k}^{(z)} + \alpha}{q_{-i,k}^{(.)} + N_z \alpha} \cdot \frac{l_{-i,k}^{(1)} + \xi_1}{l_{-i,k}^{(.)} + \xi_0 + \xi_1} \qquad (5)
$$

  In the above expression $q_{-i,k}^{z}$ is a count of number of times motif z is observed with state k when $s_i^t = 1$ removing the current observation. Similarly, and $l_{-i,k}^{1}$ is the count of $s_i = 1$ appearing with state k barring the current observation.

- Case 2: $p(s_i = 1, v_i = j)$: By our definition, this case is impossible and therefore $p(s_i = 1, v_i = j) = 0$

- Case 3: $p(s_i = 0, v_i = -1)$: Again by the definition of v_i^t, it has no meaning when $s_i = 0$. Therefore, this probability also equals zero.

- Case 4: $\forall j \in P^t, p(s_i = 0, v_i = j)$:

$$p(s_i = 0, v_i = j | c^t = k, O, S_{-i}, V_{-i}, C^{-t}, hp) \propto p(s_i = 0, v_i = j, c^t = k, S_{-i}, V_{-i}, O, C^{-t} | hp) \tag{6}$$

$$\propto p(v_i = j | s_i = 0, c^t = k, S_{-i}, V_{-i}, P^t, C^{-t}, hp)$$

$$p(s_i = 0 | o_i = z, c^t = k, S_{-i}, V_{-i}, O_{-i}, C^{-t}, hp) \tag{7}$$

$$\propto p(v_i = j | s_i = 0, P^t) p(s_i = 0 | c^t = k, S_{-i}, C^{-t}, \xi)$$

$$p(o_i = z | s_i = 0, P^t(j) = z', S_{-i}, O_{-i}, V_{-i}, \beta, \mu_0, \sigma_0^2) \tag{8}$$

$$\propto \frac{1}{N_p^t} \cdot p(s_i = 0 | c^t = k, S_{-i}, C^{-t}, \xi)$$

$$p(o_i = z | s_i = 0, P^t(j) = z', S_{-i}, O_{-i}, V_{-i}, \beta, \mu_0, \sigma_0^2) \tag{9}$$

$$\propto \frac{1}{N_p^t} \int p(s_i = 0 | \epsilon_k, c^t = k) p(\epsilon_k | S_{-i}, C^{-t}) \, \mathrm{d}\epsilon_k \cdot$$

$$\int p(o_i = z | s_i = 0, P^t(j) = z', \tau_{z'}) p(\tau_{z'} | S_{-i}, V_{-i}) \, \mathrm{d}\tau_{z'} \cdot$$

$$\int p(o_i = z | s_i = 0, P^t(j) = z', \delta_{z'}(z)) \cdot p(\delta_{z'}(z) | S_{-i}, V_{-i}) \, \mathrm{d}\delta_{z'}(z) \tag{10}$$

We move from expression 6 to expression 8 by splitting it into likelihood prior terms and simplifying using conditional independence of the variables, (i.e) v_i is conditionally independent of all other variables given s_i and P^t. Also the $\frac{1}{N_p^t}$ factor in expression 9 comes from our definition of v_i in equation 1.

There are three integrals to the r.h.s of the above equation. The first integral is similar to the second part of expression 5, but instead deals with counts of $s_i = 0$ in state k. The second integral is a Dirichlet-Multinomial distribution arising from the transition matrix. The third integral is a Gaussian-Gaussian conjugate coming from the temporal lag. Thanks to the conjugacy property, we can evaluate all the three integrals in closed form and get the following proportionality term:

$$p(s_i = 0, v_i = j | O, S_{-i}, V_{-i}, c^t = k, C^{-t}, hp) \propto \frac{1}{N_p^t} \cdot \frac{l_{-i,k}^{(0)} + \xi_0}{l_{-i,k}^{(.)} + \xi_0 + \xi_1} \cdot \frac{r_{-i,z'}^{(z)} + \beta}{r_{-i,z'}^{(.)} + N_z \cdot \beta} \cdot$$

$$\mathcal{N}\left(t - t' | \delta_{-i,z'}(z), \sigma_{z',z}^2 + \sigma^2\right) \tag{11}$$

In the above equation, $l_{-i,k}^{(0)}$ indicates the number of times $s_i = 0$ occurs with state k. $r_{-i,z'}^{(z)}$ is the number of times motif z appears after z'. $r_{-i,z'}^{(.)}$ is the total count of any event appearing after z'. $t - t'$ is the temporal lag between the occurrences $P^t(v_i) = z'$ and $o_i = z$. $\delta_{-i,z'}(z)$ and σ_n^2 are the posterior mean and variance of the temporal lag calculated from all associations of type z' and z. All the above calculations exclude the current observation. The posterior mean and variance are given by

$$\delta_{z'}(z) = \left(\frac{\mu_0}{\sigma_0^2} + \frac{D(z', z)}{\sigma^2}\right) \cdot \sigma_{z',z}^2 \tag{12}$$

$$\sigma_{z',z}^2 = \left(\frac{1}{\sigma_0^2} + \frac{r_{-i,z'}^{(z)}}{\sigma^2}\right)^{-1} \tag{13}$$

μ_0 and σ_0^2 are the prior mean and standard deviation of the lag parameter. The variance of lags between motif pairs is fixed at σ^2. $D(z', z)$ is the sum of all lag due to associations of type z' and z. For a more detailed derivation of this expression we refer to [4].

4.2. State variables

Here we derive the expression for getting the current state variable c^t. This is done by looking at all the connections coming in and out of the node c^t in the graph.

$$
\begin{aligned}
p(c^t = k|C^{-t}, S, V, O, N_o, hp) &\propto p(c^t = k, C^{-t}, S_{-i}, V, O, N_o, hp) \qquad (14)\\
&\propto \int_{\lambda_k} p(N_o^t|\lambda_k, c^t = k)p(\lambda_k|N_o^{-t}, \gamma_1, \gamma_2)\, \mathrm{d}\lambda_k \cdot \\
&\prod_{i=1}^{N_o^t} \int_{\epsilon_k} p(s_i^t|\epsilon_k, c^t = k)p(\epsilon_k|C^{-t}, S^{-t}, \xi)\, \mathrm{d}\epsilon_k \cdot \\
&\prod_{i \in \{j: s_j^t = 1\}} \int_{\theta_k} p(o_i|s_i^t = 1, \epsilon_k, c^t = k)p(\theta_k|C^{-t}, S^{-t}, O^{-t}, \alpha)\, \mathrm{d}\theta_k \\
&\int_{\pi} p(c^t = k, c^{t-1}, c^{t+1}|\pi)p(\pi|C^{-\{t-1,t,t+1\}}, S^{-t}, N_o^{-t})\, \mathrm{d}\pi \qquad (15)
\end{aligned}
$$

Here $C^{-t}, S^{-t}, O^{-t}N_o^{-t}$ refer to all the states, selector variables, and occurrences and number of occurrences except at time t. Based on the edges coming in and out of c^t we get four parts in the above equation relating to: 1) number of events, 2) number of independent and dependent events generated, 3) the motifs coming from the current state and, 4) the probability of arriving at c^t from c^{t-1} and reaching c^{t+1} from c^t. We will deal with each of them individually.

4.2.1 Occurrence Count

$$
\int_{\lambda_k} p(N_o^t|\lambda_k, c^t = k)p(\lambda_k|N_o^{-t}, \gamma_1, \gamma_2)\, \mathrm{d}\lambda_k = \int_{\lambda_k} \frac{\lambda_k^{N_o^t} \cdot e^{-\lambda_k}}{N_o^t!} \cdot \frac{\gamma_1^{\gamma_2} \lambda_k^{\sum_{i \in \{c^j = k, j \neq t\}} N_o^i + \gamma_1 - 1} \cdot e^{-\lambda_k(\gamma_2 + u)}}{\Gamma(\gamma_1)}\, \mathrm{d}\lambda_k \qquad (16)
$$

where γ_1 and γ_2 are the parameters of Gamma distribution and $u = \sum_i I(c^i = k, i \neq t)$. Making use of the Gamma-Poisson conjugacy we can write the above equation as:

$$
\int_{\lambda} p(N_o^t|\lambda_k, c^t = k)p(\lambda_k|N_o^{-t}, \gamma_1, \gamma_2)\, \mathrm{d}\lambda_k = \int_{\lambda_k} \frac{\lambda_k^{\sum_{i \in \{c^j = k, j \neq t\}} N_o^i + N_o^t + \gamma_1 - 1} e^{-\lambda_k(\gamma_2 + u + 1)} \gamma_1^{\gamma_2}}{N_o^t! \cdot \Gamma(\gamma_1)}\, \mathrm{d}\lambda_k \qquad (17)
$$

by noting that for the Gamma distribution

$$
\int_{\lambda} \mathrm{Gamma}(\lambda|\gamma_1, \gamma_2)\, \mathrm{d}\lambda = \int_{\lambda} \frac{e^{-\gamma_2 \lambda} \lambda^{\gamma_1 - 1} \gamma_2^{\gamma_1}}{\Gamma(\gamma_1)}\, \mathrm{d}\lambda = 1 \qquad (18)
$$

and by multiplying and dividing the equation by the appropriate terms and removing constant terms, we get,

$$
\int_{\lambda_k} p(N_o^t|\lambda_k, c^t = k)p(\lambda_k|N_o^{-t}, \gamma_1, \gamma_2)\, \mathrm{d}\lambda_k \propto \frac{\Gamma(\omega_1)}{N_o^t! \cdot \omega_2^{\omega_1}} \qquad (19)
$$

where, $\omega_1 = \sum_{i \in \{c^j = k, j \neq t\}} N_o^i + N_o^t + \gamma_1, \omega_2 = \gamma_2 + u + 1$

4.2.2 Motifs and Selector variables

$$\prod_{i\in\{j:s_i^t=1\}}\int_{\epsilon} p(s_i^t=1|\epsilon_k,c^t=k)p(\epsilon_k|C^{-t},S^{-t},\xi)\,\mathrm{d}\epsilon_k = \int_{\epsilon_k}\epsilon_k^{l_t^1}(1-\epsilon_k)^{l_t^0}\mathrm{Dir}(l_{-t,k}+\xi)\,\mathrm{d}\epsilon_k$$

$$= \frac{\Delta(l_t+l_{-t,k}+\xi)}{\Delta(l_{-t,k}+\xi)} \tag{20}$$

$$= \frac{\Gamma(l_t^1+l_{-t,k}^1+\xi_1)\Gamma(l_t^0+l_{-t,k}^0+\xi_0)}{\Gamma(l_t^1+l_{-t,k}^1+\xi_1+l_t^0+l_{-t,k}^0+\xi_0)}\cdot\frac{\Gamma(l_{-t,k}^0+l_{-t,k}^1+\xi_0+\xi_1)}{\Gamma(l_{-t,k}^0+\xi_0)\Gamma(l_{-t,k}^1+\xi_1)} \tag{21}$$

The above distribution is a Beta-Bernoulli distribution, with the Beta hyper-parameters $\xi=\{\xi_0,\xi_1\}$. Due to their conjugacy, we obtain the expression directly in terms of Gamma functions. where,

$$\Delta(\vec{\alpha}) = \frac{\prod_{i=1}^K\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K\alpha_i)} \tag{22}$$

In the above expression $l_t=\{l_t^0,l_t^1\}$ stands for the number of times, $s_i^t=0/1$ at the current instant t. $l_{-t,k}=\{l_{-t,k}^0,l_{-t,k}^1\}$ is the number of times $s_i^t=0/1$ from all other time instants when the state is at k except the current one.

$$\prod_{i\in\{j:s_j^t=1\}}\int_{\theta} p(o_i|s_i^t=1,\epsilon_k,c^t=k)p(\theta_k|C^{-t},S^{-t},O^{-t},\alpha)\,\mathrm{d}\theta_k = \int_{\theta_k}\prod_{z=1}^{N_z}\theta_{k,z}^{q_t^z}\mathrm{Dir}(q_{-t,k}^z+\alpha)\,\mathrm{d}\theta_k$$

$$= \frac{\Delta(q_{t,k}+q_{-t,k}+\alpha)}{\Delta(q_{-t,k}+\alpha)} \tag{23}$$

Here again we have a Dirichlet-Multinomial combination. The multinomial parameter is integrated out and the resulting expresion is in terms of Gamma functions. The count variables $q_t=\{q_t^z\}_{z=1}^{N_z}$ is the vector containing the number of times a motif z is observed at the current instant (it is either 1 or 0 in our case), and $q_{-t,k}=\{q_{-t,k}^z\}_{z=1}^{N_z}$ gives the the number of times each motif z is observed with k at all other instants except the current time t.

4.2.3 State transitions

$$\int_{\pi} p(c^t=k,c^{t-1},c^{t+1}|\pi)p(\pi|C^{-\{t-1,t,t+1\}},\varphi)\,\mathrm{d}\pi \propto \frac{n_{-i,c^{t-1}}^{c^t}+\varphi}{n_{-i,c^{t-1}}^{(\cdot)}+N_c\varphi}\cdot\frac{n_{-i,c^t}^{c^{t+1}}+I(c^{t-1}=c^t=c^{t+1})+\varphi}{n_{-i,c^t}^{(\cdot)}+I(c^{t-1}=c^t)+N_c\varphi} \tag{24}$$

The state transition is defined by the counts $n_{-i,c^{t-1}}^{c^t}$, $n_{-i,c^t}^{c^{t+1}}$, which are number of times a transition occurs from state c^{t-1} to state c^t, and from c^t to c^{t+1} after removing the current link. I is an Identity function that takes value 1, when the argument is true and 0 otherwise. $n_{-i,c^t}^{(\cdot)}$ is the the count of transitions from c^t to all other states barring the current transition. Please refer [1] for this derivation.

Putting all the four parts together we get:

$$p(c^t=k|C^{-t},S,V,O,N_o,hp) \propto \frac{\Gamma(\omega_1)}{N_o^t!\cdot\omega_2^{\omega_1}}\cdot\frac{\Gamma(l_t^1+l_{-t,k}^1+\xi_1)\Gamma(l_t^0+l_{-t,k}^0+\xi_0)}{\Gamma(l_t^1+l_{-t,k}^1+\xi_1+l_t^0+l_{-t,k}^0+\xi_0)}\cdot\frac{\Gamma(l_{-t,k}^0+l_{-t,k}^1+\xi_0+\xi_1)}{\Gamma(l_{-t,k}^0+\xi_0)\Gamma(l_{-t,k}^1+\xi_1)}\cdot$$
$$\frac{\Delta(q_{t,k}+q_{-t,k}+\alpha)}{\Delta(q_{-t,k}+\alpha)}\cdot\frac{n_{-i,c^{t-1}}^{c^t}+\varphi}{n_{-i,c^{t-1}}^{(\cdot)}+N_c\varphi}\cdot\frac{n_{-i,c^t}^{c^{t+1}}+I(c^{t-1}=c^t=c^{t+1})+\varphi}{n_{-i,c^t}^{(\cdot)}+I(c^{t-1}=c^t)+N_c\varphi} \tag{25}$$

It is important to note that, in practice calculating Delta functions $\Delta(\cdot)$ or Gamma functions $\Gamma(\cdot)$ result in overflows. This was the case while sampling the state probabilities. So we made use of log probabilities instead to calculate the state transitions. Calculations in Log probabilities are both efficient and safe.

5. Results

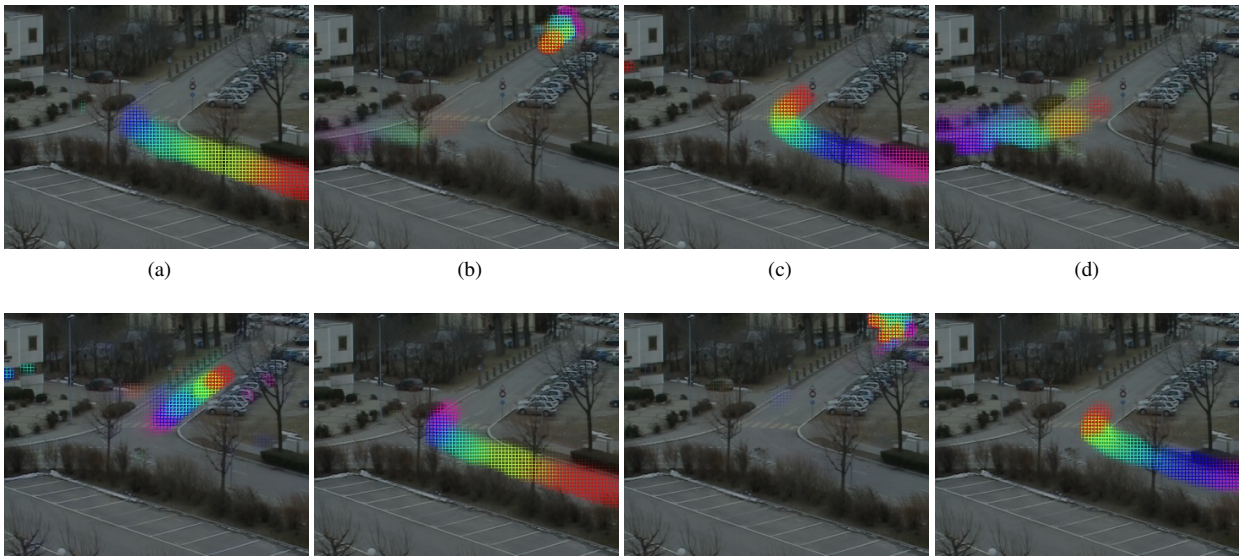
In this section we provide some more details about the inputs to the MERM model and the results obtained.

5.1. Inputs from PLSM model:

Given a video recorded from a static camera, we first generate temporal documents by first applying a PLSA step on 1 second documents made of bag-of-words of low-level location and motion features. The posterior of these recovered topics at each time instant, weighted by the number of words, is used as the input to Probabilistic Latent Sequential Motif (PLSM) algorithm. PLSM ouputs short term activities of a specified duration along with their probable times of occurrences in the video. We build a binary matrix by using these probable times of occurrences of events by a simple thresholding. Here are some examples of events and the event occurrences obtained from PLSM when applied to the four datasets we used. The binary activity occurrence matrix is the input to our MERM model.

5.1.1 Sample events from Far-field dataset

In Fig. 2 are 8 sample events out of the 15 events obtained from Far-field video by applying PLSM. These events are 5 seconds long. They represent short activities of vehicles moving in different directions in the road. The locations of the event are overlayed on the scene image with colors representing the time within the span of the event. The temporal information is color coded from violet to red. That is, locations of violet show the initial position of the event and red shows the final location of the event



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 2: 8 sample events out of the 15 events obtained from Far-field video by applying PLSM. Time within the activity is color-coded from violet to red. Locations of violet show the initial position of the event and red shows the final position of the event.

Figure 3: Activity occurrence matrix of 300 seconds from far-field data.

In the activity matrix above Fig. 3, occurrences from a set of 15 events, over 300 seconds of the far-field video is summarized. The events (as shown in the above examples) are indexed in the row and time indexed by the columns. Dark locations indicate presence of an event corresponding to that row at the time corresponding to the column.

5.1.2 Sample events from MIT dataset

Fig. 2 shows 12 sample events out of the 25 events obtained from MIT video by applying PLSM. These events are 5 seconds long. They represent short activities of various vehicular activities. Vehicles waiting (a,e) for signal, trees moving due to wind (h), and other activities in different directions (all others) in the road. The locations of the event are overlayed on the scene image with colors representing the time within the span of the event. The temporal information is color coded from violet to red. That is, locations of violet show the initial position of the event and red shows the final location of the event

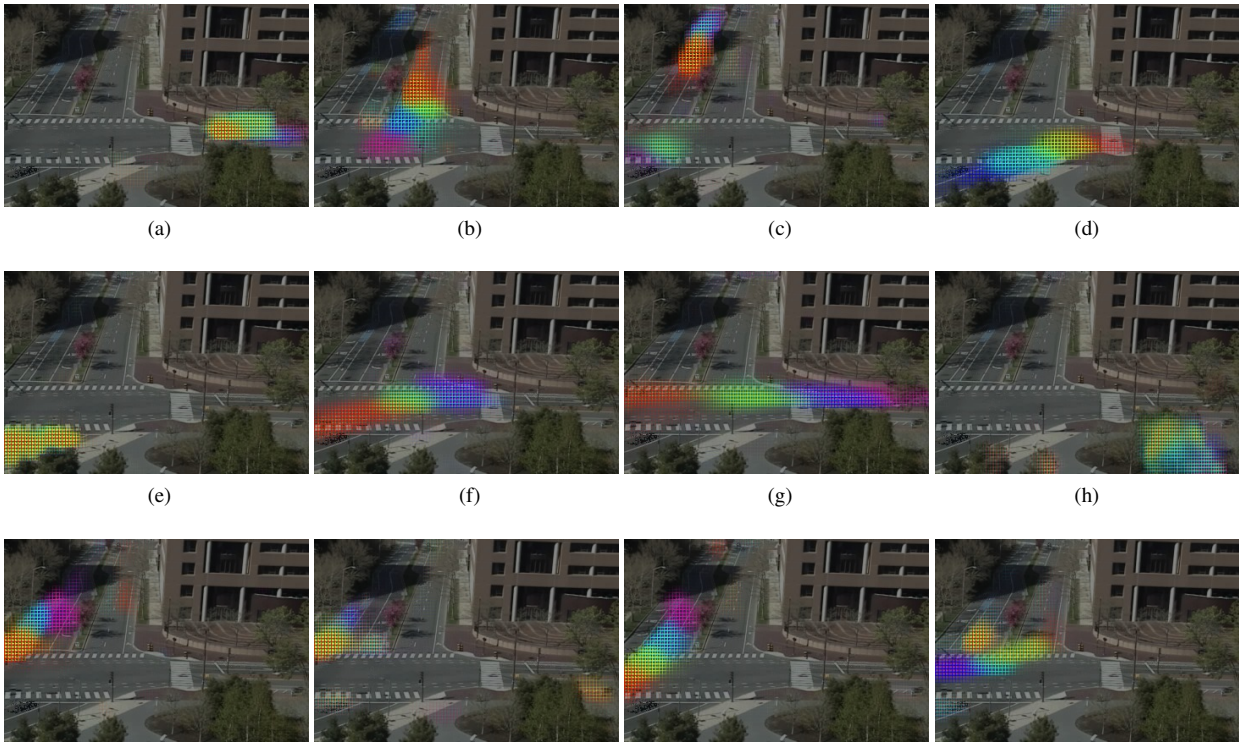


Figure 4: 12 sample events out of the 25 events obtained from MIT video by applying PLSM. Time within the activity is color-coded from violet to red. Locations of violet show the initial position of the event and red denotes the final location of the event.

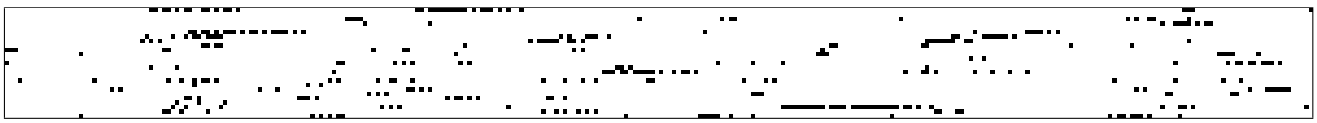


Figure 5: Activity occurrence matrix of 300 seconds from MIT data.

In the event matrix Fig. 5, occurrences from a set of 25 events, over 300 seconds of the MIT video is summarized. The events (as shown in the above examples) are indexed in the row and time indexed by the columns. Dark locations indicate presence of an event corresponding to that row at the time corresponding to the column.

5.1.3 Sample events from QMUL Junction dataset

Fig. 6 shows 5 sample events out of the 20 events obtained from UQM Junction video by applying PLSM. These events are 5 seconds long. The locations of the event are overlayed on the scene image with colors representing the time within the span of the event. The temporal information is color coded from violet to red. That is, locations of violet show the initial position of the event and red shows the final location of the event The events shown below represent movements a) from left to right, b) northwards in the left, c) southwards in the right (this is captured from UK where people drive in the left.) d) towards right and, e) from right to left.

(a) (b) (c) (d) (e)

Figure 6: 5 sample events out of the 20 events obtained from QMUL junction video by applying PLSM. Time within the activity is color-coded from violet to red. Locations of violet show the initial positions of the event and red shows the final location of the event.

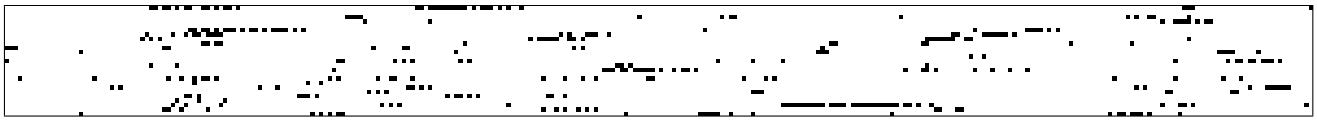


Figure 7: Activity occurrence matrix of 300 seconds from QMUL Junction data.

Fig. 7 shows an input matrix used for QMUL junction video. Occurrences from a set of 20 events, over 300 seconds of the QMUL Junction video are shown here. The events (as shown in the above examples) are indexed in the row and time indexed by the columns. Dark locations indicate presence of an event corresponding to that row at the time corresponding to the column. One might observe from this matrix that there is some periodicity in the pattern of activities occurring in the scene.

5.1.4 Sample events from ETH-Z dataset

Fig. 8 shows 8 sample events out of the 20 events obtained from ETH video by applying PLSM. These events are 5 seconds long. The locations of the event are overlayed on the scene image with colors representing the time within the span of the event. The temporal information is color coded from violet to red. That is, locations of violet show the initial position of the event and red shows the final location of the event. The events represent a variety of activities from vehicles moving along the main turn, people crossing the road (1st row, 2nd and 3rd column), a tram passing through the road (2nd row, 2nd column) etc.

Fig. 9 shows an event matrix occurrences from a set of 20 events, over 300 seconds of the ETH video. The events (as shown in the above examples) are indexed in the row and time indexed by the columns. Dark locations indicate presence of an event corresponding to that row at the time corresponding to the column.

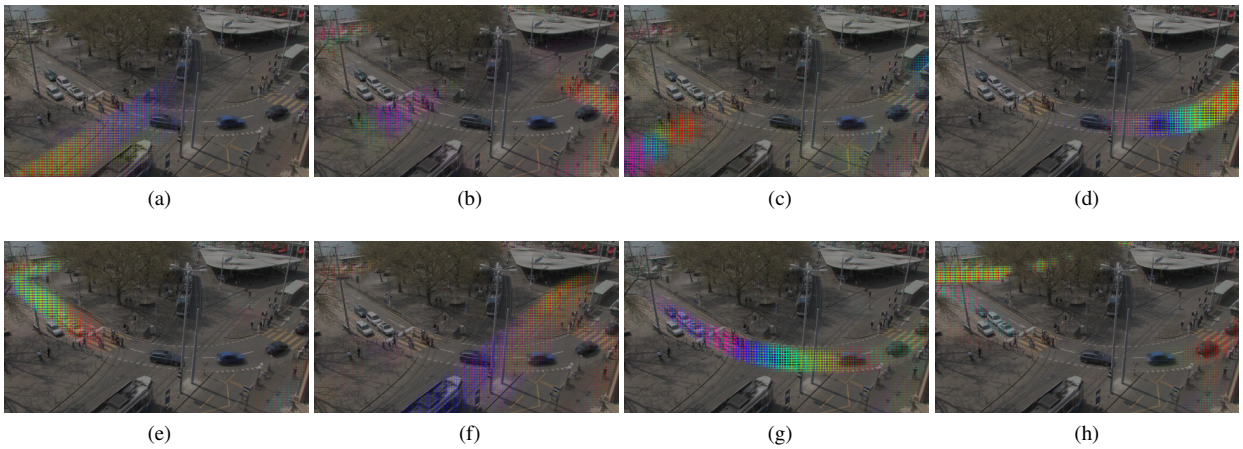


Figure 8: 8 sample events out of the 20 events obtained from ETH-Z video by applying PLSM. Time within the activity is color-coded from violet to red. Locations of violet show the initial positions of the event and red shows the final location of the event.

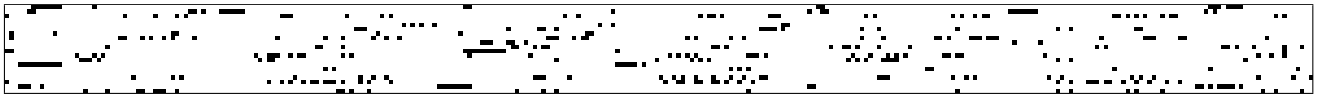


Figure 9: Activity occurrence matrix of 300 seconds from ETH-Z data.

6. Results from our MERM model

In this section we provide some of the transitions obtained by applying our MERM model in the activity matrices that were shown above.

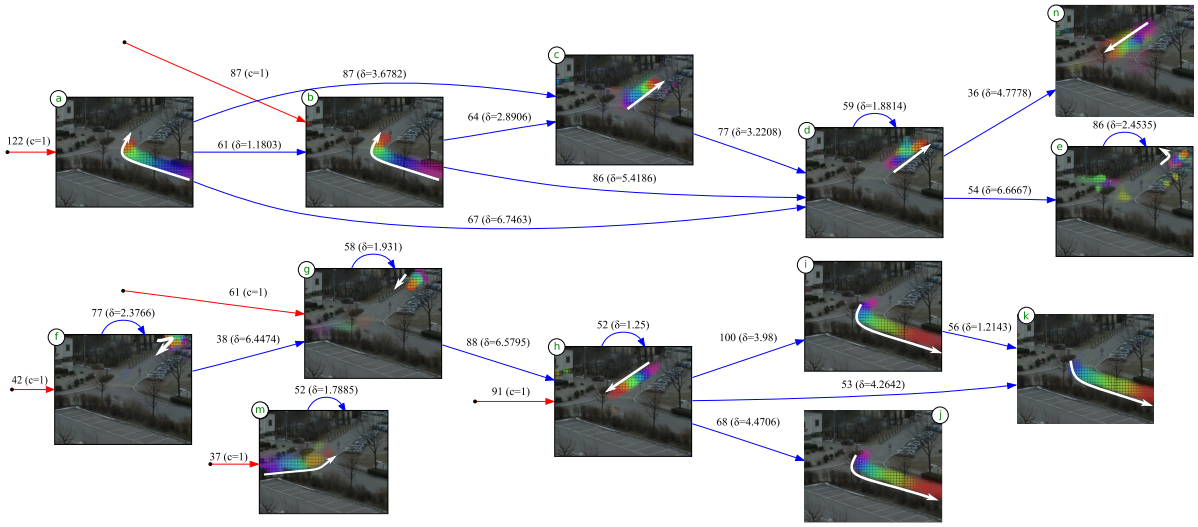


Figure 10: Event relationships from Far-field data. Low frequency edges are not shown (those with counts below 30). Blue edges: transitions with weight and lag δ. Red edges: independent activity with weight and causing state (here with only 1 state). 80% of the occurrences were dependent activities.

Fig. 10 was explained in the main paper. Fig 11 is quite similar to the results obtained from 5 seconds activities except that the activities are longer now and therefore we have lesser transitions. Here again, we obtain three main spontaneous starts

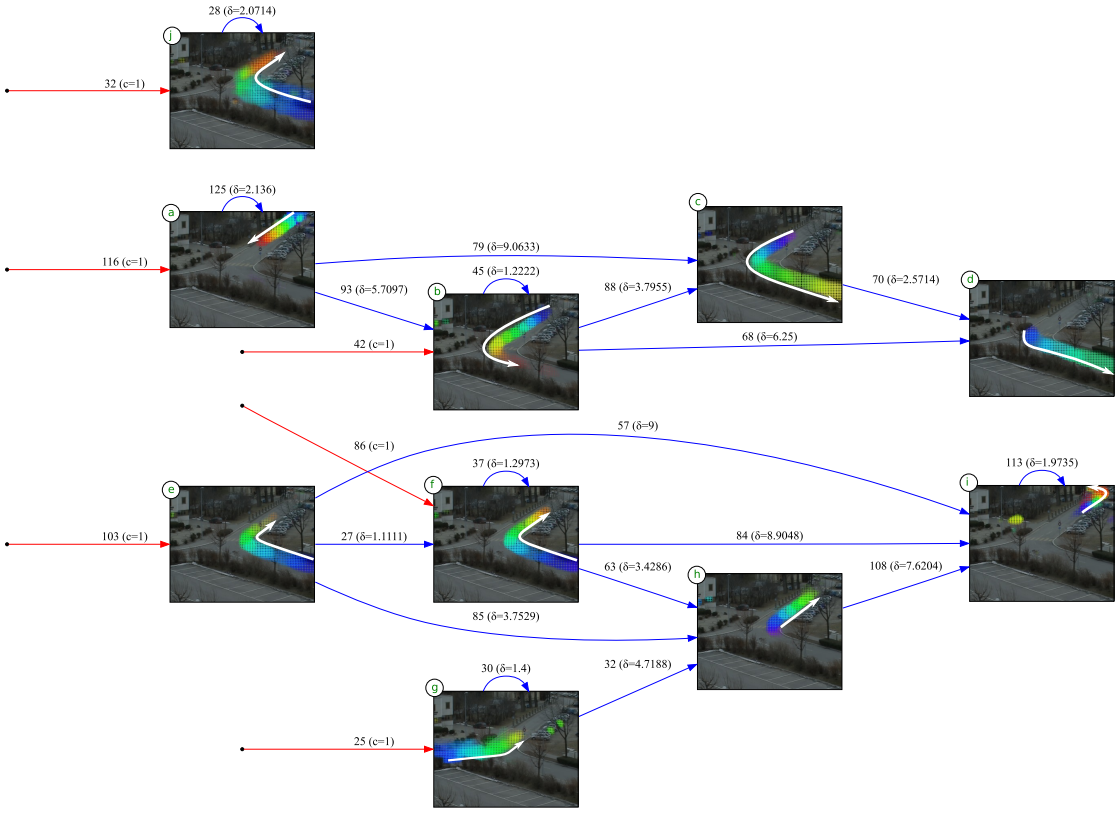


Figure 11: Event relationships from Far-field data with motifs of 10 seconds duration Low frequency edges are not shown (those with counts below 30). Blue edges: transitions with weight and lag δ. Red edges: independent activity with weight and causing state (here with only 1 state). 80% of the occurrences were dependent activities.

of events: The sequence (a,b,c,d) indicates activity starting from the top right and moving towards the bottom right. The sequence (e,f,h,i) indicates activity starting from bottom right and moving towards top right. The sequence (g,h,i) indicates vehicles starting from the bottom left and moving towards top right. Interestingly, this shares the trajectory (h,i) with the previous activity. One might notice that here, we have longer activity segments which are approximately 10 seconds long and therefore fewer transitions when compared to Fig. 10.

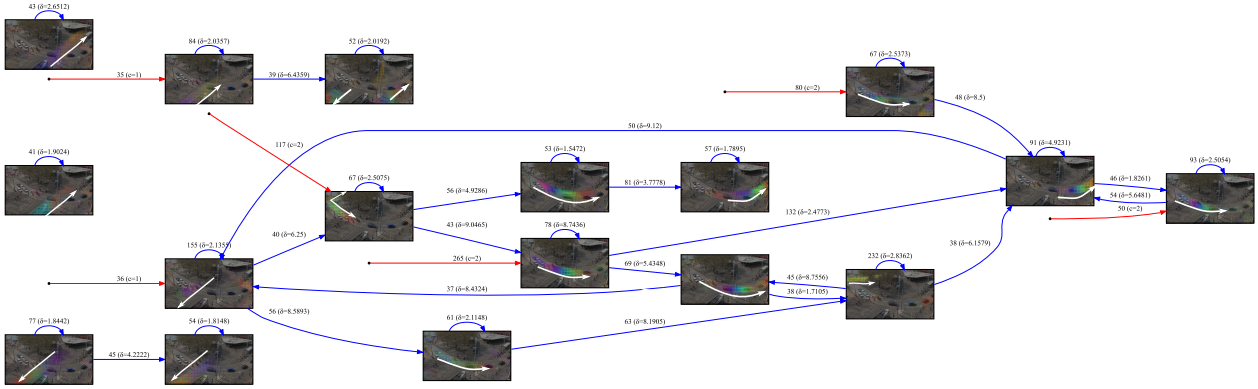
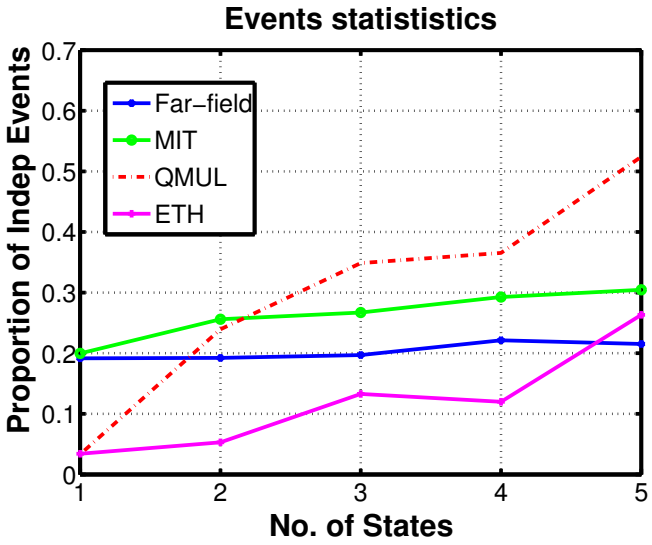


Figure 12: Event relationships from ETH-Z data with motifs of 5 seconds duration Low frequency edges are not shown (those with counts below 30). Blue edges: transitions with weight and lag δ. Red edges: independent activity with weight and causing state (here with only 1 state). 95% of the occurrences were dependent activities.

Transitions from ETH dataset is presented in Fig. 12. Experiments were run with 20 events of 5 seconds each, and asking for 2 global states from our model. We can observe that the two states capture the two dominant movements in the scene. State c1: corresponding to tram movements and people crossing in diagonal directions (botton left - top right), and state c2: corresponding to cars moving along the curved road (from top-left to top-right) in the scene. In both the states, we observe that many of the activities have self loops, indicating that the same activity is repeated multiple times one after the other. This is typically true when multiple groups of people cross the road or when cars follow each other through the curved path.

7. Ratio between Independent and dependent events

Here we show the proportion of independent activities obtained from each dataset by varying the number of states. It shows that Far-field datasets has a very stable number of independent activities. In case of Far-field data, one cannot categorize the activities into distinct states and therefor the number of independent events are invariant to the number of states. In case of MIT, the number of independent events seem to saturate when number of states is 3 and 4. In case of Junction and ETH, there are predictable patterns occurring in the scene which are more explained by the states as their number is increased. This demonstrates the typical competition in the model to explain an activity by either the global scene states or by the past events. When there are very predictable recurrent events, the model associates them to the global scene states rather than to past events as their number of states is increased.



References

[1] Integrating topics and syntax. In *Advances in Neural Information Processing Systems*, 2004. 6

[2] D. M. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, pages 993–1022, 2003. 3

[3] G. Heinrich. Parameter estimation for text analysis. Technical report, 2004. 3

[4] K. P. Murphy. Bayesian statistics: a concise introduction. Technical report, 2007. 4