Supplementary material (inference and sampling equations) for "*Extracting and Locating Temporal Motifs in Video Scenes Using a Hierarchical Non Parametric Bayesian Model*"

Rémi Emonet[1] Jagannadan Varadarajan[1,2] Jean-Marc Odobez[1,2]

[1]: Idiap Research Institute – CH-1920 Martigny, Switzerland
[2]: École Polytechnique Fédéral de Lausanne – CH-1015, Lausanne, Switzerland
{remonet,vjagann,odobez}@idiap.ch

This material details the sampling equations used for inference in our temporal topic model. To make it self-contained, we first recall the generative model equations, then provide additional notations, and finally present the actual inference equations.

1. Generative model and notations

The graphical model is reproduced in Fig. 1 It corresponds to the following set of equations:

$$
\begin{aligned}
H &= Dir(\eta) && (1)\\
\beta^M &\sim GEM(\gamma) \qquad \phi_l \sim H && (2)\\
\beta^o_j &\sim GEM(\alpha) && (3)\\
ost_{jo} &\sim U_j \text{ and } k_{jo} \sim \beta^M && (4)\\
o_{ji} &\sim \beta^o_j && (5)\\
z_{ji} &= k_{jo_{ji}} \text{ and } st_{ji} = ost_{jo_{ji}} && (6)\\
(rt_{ji}, w_{ji}) &\sim \text{Mult}(\phi_{z_{ji}}) && (7)\\
at_{ji} &= st_{ji} + rt_{ji} && (8)
\end{aligned}
$$

In this document, we will also use the following additional notations:

- $O = \{o_{ji}\}$ denotes the set of all occurrence indices associated with observations;
- $W = \{(w_{ji}, at_{ji})\}$ denotes the set of all observations, each consisting of a (word, absolute time stamp) pair;
- $K = \{k_{jo}\}$ denotes the set of all motif[1] indices associated with all existing occurrences;
- $OST = \{ost_{jo}\}$ denotes the set of starting times associated with all currently existing occurrences.
- $O^{-ji} = O - \{o_{ji}\}$ will be used to denote removing the element ji from the corresponding set. Similarly, K^{-jo} denotes $K - \{k_{jo}\}$, i.e. the set of occurrence topics except that of the joth occurrence;
- $N_{\mathcal{V}}$ will denote the size of the vocabulary $\mathcal{V}$ (i.e. the number of different words w), while L_M denotes the maximum number of relative time steps in a motif;
- we will use the notation N to denotes counts. For instance, $N_{obs}(w, rt, l) = |\{(j'i')|w_{j'i'} = w, rt_{j'i'} = rt, z_{j'i'} = l\}|$ denotes the number of observations whose associated topic and relative time are l and rt. Similarly, we note: $N_{obs}(j, o) = |\{i|o_{ji} = o\}|$, the size of the set of observations associated with the occurrence o in document j, and $N_{occ}(j, l) =$

[1] In this document, we will use the words motifs or topic in an interchangeable manner.

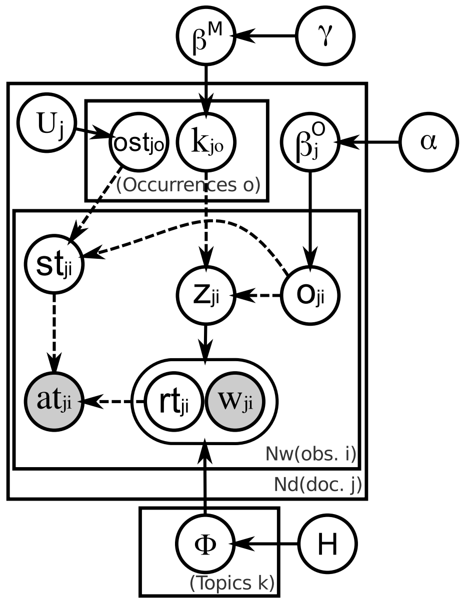

Figure 1. Proposed model using the stick-breaking convention. Dashed arrows represents deterministic relations (conditional distributions are a Dirac).

$|\{o \mid k_{jo} = l\}|$ the size of the set of occurrences in document j which are associated with topic l. As with sets, we will use the notation $-$ to remove an element from the count, e.g.
$N_{obs}^{-ji}(w, rt, l) = |\{(j'i'), (j'i') \neq (ji) | w_{j'i'} = w, rt_{j'i'} = rt, z_{j'i'} = l\}|$.

2. Inference and sampling equations

The inference is conducted using a Gibbs sampling approach that is run until a large and fix number of iterations is performed. Given our model, the sampling will be conducted in turn over the following set of variables:

- o_{ji}: the occurrence index associated with a given observation;
- k_{jo}: the motif k associated with a given occurrence jo (occurrence o in document j);
- ost_{jo}: the starting time associated with the occurrence o of document j.

Other variables (in particular the β^M, β^o of the GEM processes, as well as the parameters of the multinomial distribution ϕ_k) will be integrated out. Below we detail the posterior probability associated with each variable. To simplify notations, in the conditional, we will not make reference to the variables which are not sampled, i.e. the set of observations W, the concentration α and γ, and the prior parameters of the motif η.

2.1. Sampling o_{ji} for some given j and i

To sample this variable, we have to compute its posterior distribution given all other variables. It is given by:

$$p_{draw}(o_{ji} = o) \quad \triangleq \quad p(o_{ji} = o | O^{-ji}, K, OST) \quad \propto p(w_{ji}, at_{ji} | o_{ji} = o, O^{-ji}, K, OST) p(o_{ji} = o | O^{-ji}) \tag{9}$$

The candidates values o for o_{ji} are of two kinds: the set of currently existing occurrences indices in document j, and an additional index u representing the case where the observation is associated with a new occurrence. Below we detail the posterior probability computation for the two cases.

2.1.1 Posterior of an existing occurrence o.

In this case, the first term in the right hand side of Eq. 9 can be computed as:

$$\begin{aligned} p(w_{ji}, at_{ji}|o_{ji}=o, O^{-ji}, K, OST) &= p(w_{ji}, at_{ji}|z_{ji}=k_{jo}, st_{ji}=ost_{jo}, O^{-ji}) \\ &= p(w_{ji}, rt_{ji}|z_{ji}=k_{jo}, O^{-ji}) \qquad \text{given that } rt_{ji}=at_{ji}-st_{j,i} \\ &= \frac{N_{obs}^{-ji}(w_{ji}, rt_{ji}, z_{ji}) + \eta(w_{ji}, rt_{ji})}{\sum_{w',rt'} \Big(N_{obs}^{-ji}(w', rt', z_{ji}) + \eta(w', rt')\Big)} \end{aligned} \tag{10}$$

where the last equation is obtained by exploiting our use of a Dirichlet conjugate prior to integrate out the multinomial parameters of the motif z_{ji}. The second term in the right hand side of Eq. 9 (the probability of drawing an existing occurrence) is given by:

$$p(o_{ji}=o|O^{-ji}) = \frac{N_{obs}^{-ji}(j,o)}{\sum_{o'} N_{obs}^{-ji}(j,o') + \alpha} = \frac{N_{obs}^{-ji}(j,o)}{N_j - 1 + \alpha} \qquad \text{from the Chinese restaurant process (Dirichlet Process)} \tag{11}$$

where N_j denotes the number of observations in document j.

2.1.2 Posterior of a new occurrence $o = u$

In this case, the second term in the right hand side is directly given by the Chinese restaurant process of the occurrence level:

$$p(o_{ji}=u|O^{-ji}) = \frac{\alpha}{N_j - 1 + \alpha}. \tag{12}$$

The first term is however more difficult to compute. Indeed, the new occurrence index u can be associated with any newly drawn possible occurrence. In other words, we need to marginalize (i.e. integrate) the likelihood over all possible draw of a new occurrence (k_{ju}, ost_{ju}). Thus we have:

$$p(w_{ji}, at_{ji}|o_{ji}=u, O^{-ji}, K, OST) \tag{13}$$

$$\begin{aligned} &= \mathrm{E}_{(k_{ju}, ost_{ju})|K,OST}\left[p(w_{ji}, at_{ji}|k_{ju}, ost_{ju})\right] \\ &= \sum_{k_{ju}} \sum_{ost_{ju}} p(w_{ji}, at_{ji}|ost_{ju}, k_{ju}) p(k_{ju}, ost_{ju}|K, U_j) \\ &= \sum_{k_{ju}} \sum_{ost_{ju}} p(w_{ji}, at_{ji}|ost_{ju}, k_{ju}) p(k_{ju}|K) p(ost_{ju}|U_j) \qquad k_{ju} \text{ and } ost_{ju} \text{ are drawn independently} \\ &= \sum_{k_{ju}} \sum_{ost_{ju}} p(w_{ji}, at_{ji}|ost_{ju}, k_{ju}) p(k_{ju}|K) \frac{1}{L_j} \qquad \text{uniform prior } U_j \\ &= \sum_{k_{ju}} \sum_{rt_u} p(w_{ji}, rt_u|k_{ju}) p(k_{ju}|K) \frac{1}{L_j} \qquad \text{using the change of variable } rt_u = at_{ji} - ost_{ju} \\ &= \frac{1}{L_j} \sum_{k_{ju}} \sum_{rt_u} p(k_{ju}|K) p(w_{ji}, rt_u|k_{ju}) \end{aligned} \tag{14}$$

where L_j denotes the number of possible starting times for an occurrence. In the last expression we need to distinguish two cases for k_{ju} when computing the two terms ($p(k_{ju}|K)$ and $p(w_{ji}, rt_u|k_{ju})$) in the summation: either k_{ju} is a motif that already exists (i.e. it is already associated with at least one occurrence in all documents) or it is a completely new motif.

According to the Chinese restaurant process at the motif level, the probability of drawing an existing motif m is proportional to the number $N_{occ}(m) = \sum_j N_{occ}(j,m)$ of occurrences in all the documents that are associated with that motif, whereas the probability of drawing a new motif is proportional to γ. In equations:

$$p(k_{ju} = m|K) = \begin{cases} \dfrac{N_{occ}(m)}{\sum\limits_{m'} N_{occ}(m') + \gamma} & \text{if } m \text{ is an existing motif } m \\ \dfrac{\gamma}{\sum\limits_{m'} N_{occ}(m') + \gamma} & \text{if } m = m_{new} \text{ is a new motif} \end{cases} \tag{15}$$

Similarly, for the computation of the observation likelihood term, we have the two cases. For an existing motif, it can be computed as was already presented in Eq.10, i.e. we have:

$$p(w_{ji}, rt_u | k_{ju} = m) \quad = \quad \frac{N_{obs}^{-ji}(w_{ji}, rt_u, z_{ji} = m) + \eta(w_{ji}, rt_u)}{\sum\limits_{w', rt'} \left(N_{obs}^{-ji}(w', rt', z_{ji} = m) + \eta(w', rt') \right)} \text{ if } m \text{ is an existing motif.} \tag{16}$$

When the motif is new, the same applies but as there are no observation yet associated with that motif, the likelihood corresponds to the prior density, i.e.:

$$p(w_{ji}, rt_u | k_{ju} = m) = \frac{\eta(w_{ji}, rt_u)}{\sum\limits_{w', rt'} \eta(w', rt')} \text{ if } m \text{ is a new motif.} \tag{17}$$

2.1.3 Summary

To sample o_{ji}, we compute the posterior from Eq. 9 and all equations that follows. We then sample the new occurrence from the posterior. If an existing occurrence is selected, we just need to modify the index for the given observation (and do the bookkeeping for the table N_{obs}).

If a new occurrence o_{new} is sampled, then the elements $(ost_{jo_{new}}, k_{jo_{new}})$ associated with that occurrence need to be sampled as well. This is done by sampling from the mixture on (k_{ju}, rt_u) in Eq. 14[2].

2.2. Sampling k_{jo}

When updating an occurrence, we need to sample both the associated motif k_{jo} and the occurrence starting time ost_{jo}. Currently, we are updating both variable separately. Although it can be seen as sub-optimal (since the motif and their associated starting occurrences can be strongly coupled through their associated observations), in practice we did not observe this as a major problem during learning.

As the reassignment of k_{jo} to some motif m affects all observations associated with this occurrence o, the computation of the posterior for this variable involves the computations of the likelihood for all these observations. Denoting by $I(o)$ the set of indices in document j such $o_{ji} = o$, the posterior for k_{jo} can be written as:

$$p(k_{jo} = m | O, K^{-jo}, OST) = p(k_{jo} = m | K^{-jo}) \prod_{i \in I(o)} p(w_{ji}, at_{ji} | k_{jo} = m, ost_{jo}) \tag{18}$$

where the first term can be computed as in Eq. 15 (i.e. the probability is proportional to $N_{occ}^{-jo}(m)$ for an existing motif, and γ for a new motif), and the likelihood term can be computed similarly to Eq. 10, but excluding all observations $ji, i \in I(o)$ when doing the counts:

$$p(w_{ji}, at_{ji} | k_{jo} = m, ost_{jo}) = \frac{N_{obs}^{-jI(o)}(w_{ji}, rt_{ji} = at_{ji} - ost_{jo}, z_{ji} = m) + \eta(w_{ji}, rt_{ji} = at_{ji} - ost_{jo})}{\sum\limits_{w', rt'} \left(N_{obs}^{-jI(o)}(w', rt', z_{ji} = m) + \eta(w', rt') \right)}. \tag{19}$$

[2] Note that given the sampled $rt_u^\star$, the $ost_{jo_{new}}$ can be set as $at_{ji} - rt_u^\star$.

2.3. Sampling ost_{jo}

The sampling of ost_{jo} is conducted as with k_{jo}, but involves only the observation likelihood since the prior on ost values is uniform. Thus, we have:

$$p(ost_{jo} = ost|O, OST^{-jo}) \propto \prod_{i \in I(o)} p(w_{ji}, at_{ji}|k_{jo}, ost_{jo} = ost) \tag{20}$$

and $p(w_{ji}, at_{ji}|k_{jo}, ost_{jo} = ost)$ is given by Eq. 19.

In addition, and in order to favor the alignment of learned motifs so that they exhibit a significant amount of activity right after their starting time (See paper section 4.2 and Fig. 5 in paper), we also propose to modify all occurrence starting times of individual motifs jointly, by sampling the same temporal increment that will apply to all of them. The posterior for this increment for a given motif m is then given by the likelihood of all observations currently associated with this motif:

$$p(Inc = inc|O, OST, K, m) \propto \prod_{ji} \prod_{s.t. k_{jo_{ji}} = m} p(w_{ji}, at_{ji}|m, st_{ji} = ost_{jo_{ji}} + inc). \tag{21}$$

This likelihood can be computed as done previously. However, notice that since all observations associated with the motif m are removed, this likelihood function is simply the prior density for the observations and we have:

$$p(w_{ji}, at_{ji}|m, st_{ji} = ost_{jo_{ji}} + inc) \propto \eta(w_{ji}, rt_{ji} = at_{ji} - st_{ji}) = \eta(w_{ji}, rt_{ji} = at_{ji} - ost_{jo_{ji}} - inc) \tag{22}$$

When the increment is sampled, the starting times of all occurrences of motif m are modified accordingly. The consequence is that the relative time rt_{ji} of all observations is automatically shifted in the opposite direction (cf Eq.6 and 8), which ultimately modifies the count table N_{obs} associated with the motif m by the same temporal shift.

In practice, the main effect of this sampling is to produce count tables for the motifs whose marginal w.r.t. the relative temporal occurrence of words is matching the prior we have set (See paper section 4.2 and Fig. 5 in paper).

3. Acknowledgements

The authors gratefully acknowledge the financial support from the Swiss National Science Foundation (Project: FNS-198,HAI) and from the 7th framework program of the European Union integrated project VANAHEIM (248907) under which this work was carried out.