Dynamic Bayesian Poisson Tensor Factorization

Aaron Schein — UMass Amherst
Mingyuan Zhou — Univ. of Texas Austin
John Paisley — Columbia Univ.
David M. Blei — Columbia Univ.
Hanna Wallach — Microsoft Research

Overview

- $\mathbf{Y}_t$ is an M-way count tensor of size $D_1 \times \cdots \times D_M$.
- We observe it evolve over T time steps: $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$.
- For example, typed interactions between country actors:

$$\mathbf{Y}_t \in \mathbb{N}^{N \times N \times A} \quad (N \text{ country actors, } A \text{ different action types}),$$
$$y_{ijat} \equiv \# \text{ actions of type } a \text{ taken by country } i \text{ towards } j \text{ at time } t.$$

- Goal: Infer latent components with temporally-smooth intensity.
- We fit the dynamic model and non-dynamic baseline model to GDELT which contains dyadic events between country actors that are automatically extracted from Internet news archives, e.g.:

$$\langle \text{Turkey, Syria, } \textit{Fight}\text{, 12/25/2014} \rangle$$

Dec. 25, 2014: "Turkish jets bombed targets in Syria."

Model

Observed counts are drawn from a Poisson distribution:

$$y_{\mathbf{i}t} \sim \text{Pois}\left(\sum_{k=1}^{\infty} \lambda_k \, \phi^{(1)}_{k\mathbf{i}_1} \cdots \phi^{(M)}_{k\mathbf{i}_M} \theta_{kt}\right),$$

where $\mathbf{i} \equiv (\mathbf{i}_1, \cdots, \mathbf{i}_M)$ is a multi-index into $\mathbf{Y}_t$.

By Poisson additivity, $y_{\mathbf{i}t}$ can also be expressed as the sum of *latent sources*:

$$y_{\mathbf{i}t} = \sum_{k=1}^{K} y_{\mathbf{i}tk}, \qquad y_{\mathbf{i}tk} \sim \text{Pois}\left(\lambda_k \, \phi^{(1)}_{k\mathbf{i}_1} \cdots \phi^{(M)}_{k\mathbf{i}_M} \theta_{kt}\right).$$

Weights λ_k are drawn from a Gamma process on $\mathbb{R}_+ \times \Omega$:

$$G \equiv \sum_{k=1}^{\infty} \lambda_k \delta_{\omega_k} \sim \text{GaP}(c, G_0),$$

where c is the concentration parameter, G_0 is the base measure on $\mathbb{R}_+ \times \Omega$, and an atom $\omega_k \in \Omega$ consists of $M{+}1$ factor vectors:

$$\omega_k \equiv (\boldsymbol{\phi}^{(1)}_k, \ldots, \boldsymbol{\phi}^{(M)}_k, \boldsymbol{\theta}_k).$$

In practice we set K and assume the weights are drawn:

$$\lambda_k \sim \text{Gamma}\left(\frac{\gamma_0}{K}, 1/c\right),$$

which approaches the Gamma process as $K \to \infty$.

The non-dynamic factor vectors are drawn from Dirichlet distributions:

$$\boldsymbol{\phi}^{(m)}_k \sim \text{Dir}\left(\boldsymbol{\eta}^{(m)}\right), \quad \text{where } \boldsymbol{\eta}^{(m)} \equiv (\eta^{(m)}_1, \cdots, \eta^{(m)}_{D_m}).$$

The dynamic factors are drawn from a Gamma Markov chain:

$$\theta_{k(t+1)} \sim \text{Gamma}\left(\theta_{kt}, \; 1/c_k\right),$$

where each is a Gamma r.v. with its *shape* equal to the previous factor.

The inverse scale c_k parameters are drawn from a Gamma distribution:

$$c_k \sim \text{Gamma}\left(e_0, 1/f_0\right),$$

along with the Gamma process concentration and mass parameter $\gamma_0 \equiv G_0(\Omega)$:

$$c \sim \text{Gamma}\left(e_0, 1/f_0\right) \qquad \gamma_0 \sim \text{Gamma}\left(e_0, 1/f_0\right).$$

Key properties

Three different ways of generating a Negative Binomial random variable y:

$$y \sim \text{NB}(r, p) \qquad \lambda \sim \text{Gamma}\left(r, \tfrac{p}{1-p}\right), \; y \sim \text{Pois}(\lambda) \qquad l \sim \text{Pois}\big(-r\ln(1-p)\big), \; y \sim \text{SumLog}(l, p)$$

A SumLog random variable with parameters $l \in \mathbb{N}$ and $p \in (0,1)$ is defined as the sum of iid logarithmic random variables:

$$y \sim \text{SumLog}(l, p) \qquad y = \sum_{j=1}^{l} u_j, \quad u_j \overset{\text{iid}}{\sim} \text{Log}(p).$$

Conditioned on y and r, the conditional posterior over l is a Chinese Restaurant Table distribution $P(l|y,r) = \text{CRT}(l; y, r)$ which can be defined as a sum of Bernoulli random variables:

$$l \sim \text{CRT}(y, r), \qquad l = \sum_{j=1}^{y} z_j, \quad z_j \sim \text{Bern}\left(\tfrac{r}{j-1+r}\right).$$

The joint distribution of l and y is a bivariate count distribution called the Poisson-logarithmic distribution $P(y, l|r, p) = \text{PoisLog}(y, l; r, p)$ that can be constructed differently under two different factorizations:

$$P(y, l|r, p) \;\; = P(l|y, r)\, P(y|r, p) \qquad = P(y|l, p)\, P(l|r, p)$$
$$\text{PoisLog}(y, l|r, p) \;\; = \text{CRT}(l; y, r)\, \text{NB}(y; r, p) \qquad = \text{SumLog}(y; l, p)\, \text{Pois}\big(l; -r\ln(1-p)\big)$$

MCMC inference

Conditioned on $y_{\mathbf{i}t}$ and all other latent variables, the conditional posterior over latent sources is Multinomial where we define $\zeta_{\mathbf{i}tk} \equiv \lambda_k \, \phi^{(1)}_{k\mathbf{i}_1} \cdots \phi^{(M)}_{k\mathbf{i}_M} \theta_{kt}$:

$$(y_{\mathbf{i}t1}, \cdots, y_{\mathbf{i}tK} | -) \sim \text{Multi}\Big(y_{\mathbf{i}t}, \; \big(\tfrac{\zeta_{\mathbf{i}t1}}{\sum_{k=1}^{K} \zeta_{\mathbf{i}tk}}, \cdots, \tfrac{\zeta_{\mathbf{i}tK}}{\sum_{k=1}^{K} \zeta_{\mathbf{i}tk}}\big)\Big).$$

By Dirichlet-Multinomial conjugacy, the conditional posterior over non-dynamic factor vectors is:

$$(\boldsymbol{\phi}^{(m)}_k | -) \sim \text{Dir}\Big(\eta^{(m)}_1 + y^{(m)}_{1k}, \cdots, \eta^{(m)}_{D_m} + y^{(m)}_{D_m k}\Big), \qquad y^{(m)}_{dk} \equiv \sum_{\mathbf{i}:\, \mathbf{i}_m = d} \; \sum_{t=1}^{T} y_{\mathbf{i}tk}.$$

A forwards-backwards algorithm developed by Acharya, Ghosh, & Zhou (AISTATS 2015) for conditionally conjugate inference of the dynamic factors involves sampling auxiliary latent counts:

Backwards:

$$(l_{tk}|-) \sim \text{CRT}\big(y_{tk} + l_{(t+1)k}, \theta_{k(t-1)}\big), \qquad p_{tk} := \frac{\lambda_k - \ln(1 - p_{(t+1)k})}{c_k + \lambda_k - \ln(1 - p_{(t+1)k})}$$

Forwards:

$$(\theta_{kt}|-) \sim \text{Gamma}\big(\theta_{k(t-1)} + y_{tk} + l_{(t+1)k}, \; \tfrac{1 - p_{tk}}{c_k}\big)$$

The key idea is that the final factor θ_{kT} can be marginalized out, yielding:

$$y_{Tk} \sim \text{NB}\big(\theta_{k(T-1)}, p_{Tk}\big), \qquad p_{Tk} \equiv \tfrac{\lambda_k}{c_k + \lambda_k}.$$

The negative binomial can then be augmented with an auxiliary count $l_k \sim \text{CRT}(y_{Tk}, \theta_{k(T-1)})$ resulting in the CRT-NB construction of the PoisLog:

$$(l_{Tk}, y_{Tk}) \sim \text{CRT}(l_{Tk}; y_{Tk}, \theta_{k(T-1)})\, \text{NB}(y_{Tk}; \theta_{k(T-1)}, p_{Tk})$$

which can then be re-expressed as the SumLog-Poisson construction:

$$(y_{Tk}, l_{Tk}) \sim \text{SumLog}(y_{Tk}; l_{Tk}, p_{Tk})\, \text{Pois}\big(l_{Tk}; -\theta_{k(T-1)} \ln(1 - p_{Tk})\big)$$

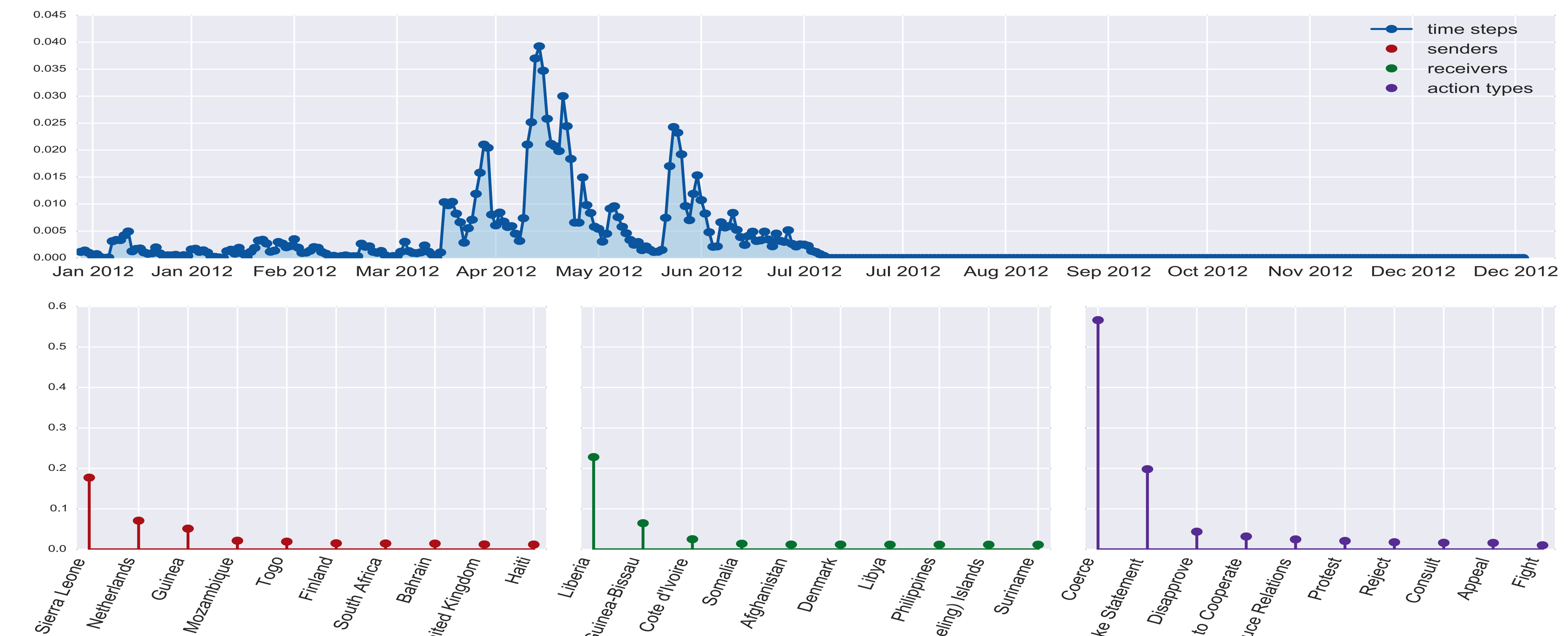
Thus l_{Tk} can be sampled as an auxiliary CRT variable and then considered backwards information to $\theta_{k(T-1)}$. For more details see Acharya et al.

The Gamma process mass parameter γ_0 can be similarly sampled:

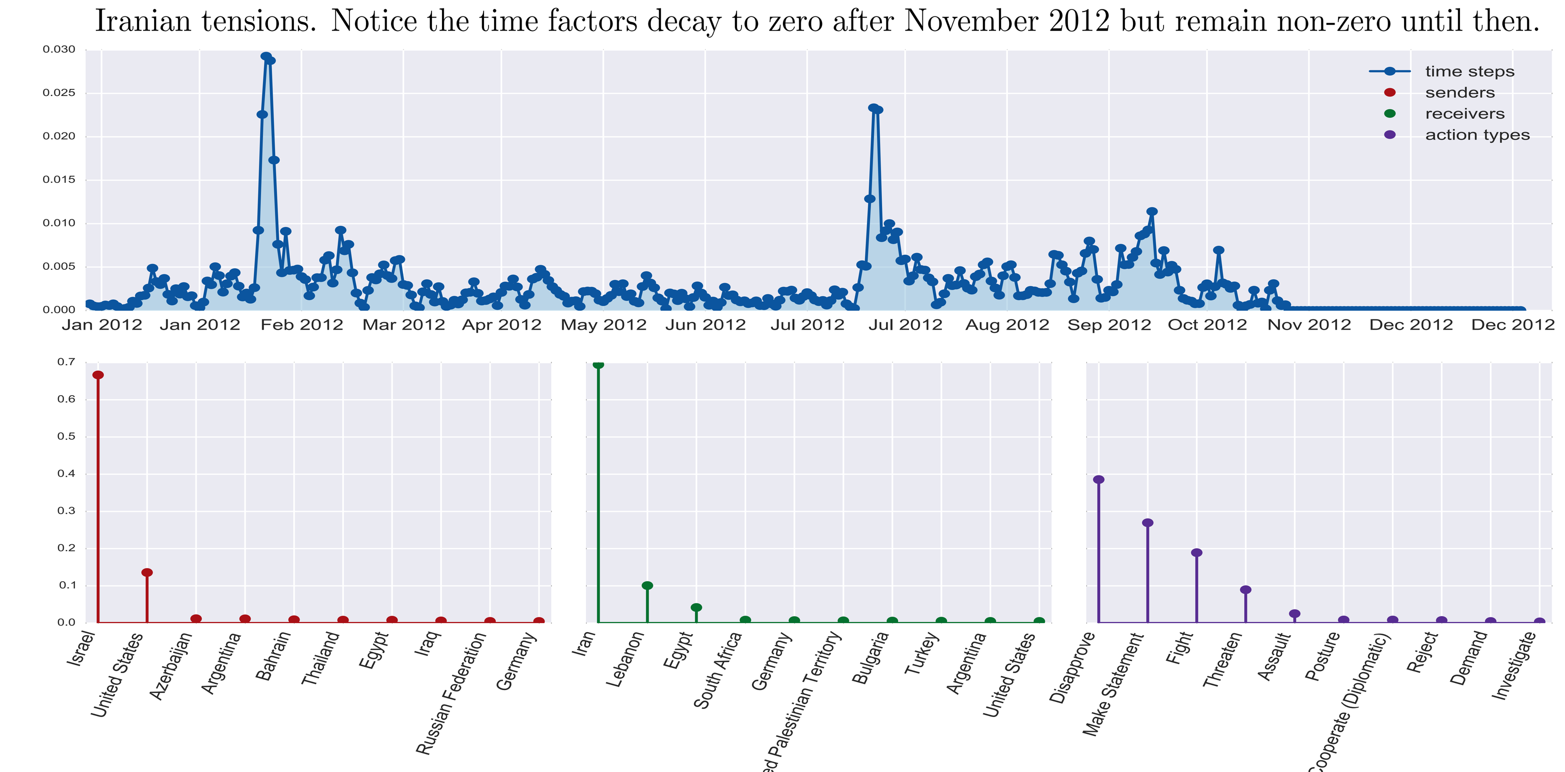
$$(l_k|-) \sim \text{CRT}\big(y_k, \; \tfrac{\gamma_0}{K}\big)$$
$$(\gamma_0|-) \sim \text{Gamma}\Big(e_0 + \textstyle\sum_{k=1}^{K} l_k, \; \big(f_0 - \tfrac{1}{K} \sum_{k=1}^{K} \ln(1 - \tfrac{\sum_t \theta_{tk}}{c + \sum_t \theta_{tk}})\big)\Big)$$

Example results on GDELT

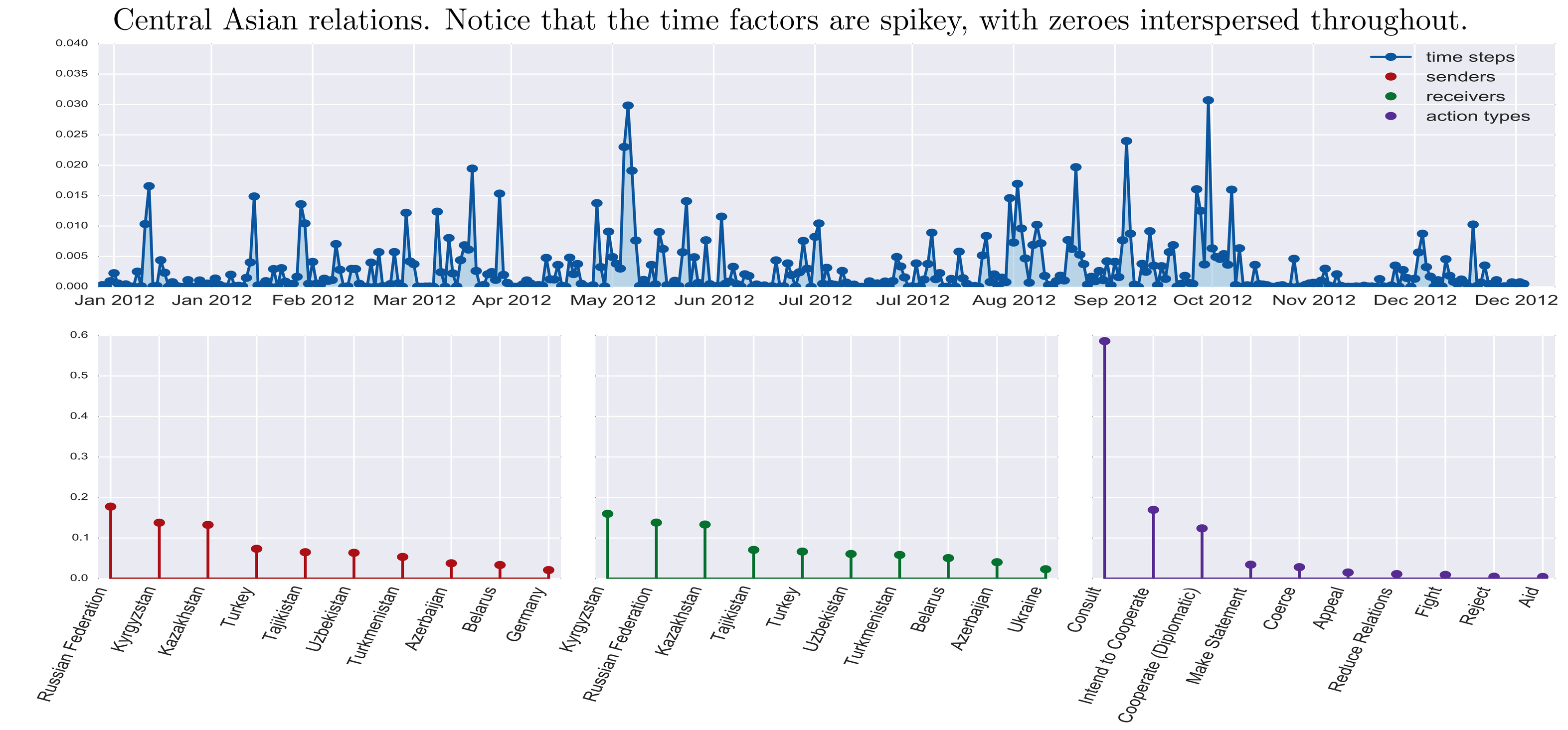
A component inferred by the dynamic model. The top actors and time factors suggest this corresponds to the conviction of Charles Taylor, former president of Liberia, in April 2012, for war crimes during the Sierra Leone civil war. Notice the time factors decay to zero after June 2012 but remain non-zero until then.



A component inferred by the dynamic model. Top actors and time factors suggest this corresponds to Israeli-Iranian tensions. Notice the time factors decay to zero after November 2012 but remain non-zero until then.



A component inferred by the non-dynamic model. Top actors and time factors suggest this corresponds to Central Asian relations. Notice that the time factors are spikey, with zeroes interspersed throughout.



Left: Lag-1 autocorrelation of the time series for (i, j) country-country pairs. Countries are sorted by overall activity in the dataset; the upper-left portion of the matrix contains more active country-country pairs. The time series for the top 25 countries show positive autocorrelation while the majority of time series exhibit zero or negative autocorrelation. This suggests that only a small (but important) subset of the data motivates a dynamic model. **Middle:** Correlation coefficients between time steps in the latent time step factors inferred by the dynamic model. Notice that the dynamic model infers strong covariance between close time steps and dampens covariance at distant time steps. **Right:** The same plot for factors inferred from the nondynamic baseline. Notice that this model infers unstructured covariance at long time lags.