# Detecting and Characterizing Events: Appendices

**Allison J. B. Chaney**
Princeton University
achaney@cs.princeton.edu

**Hanna Wallach**
Microsoft Research
wallach@microsoft.com

**Matthew Connelly**
Columbia University
mjc96@columbia.edu

**David M. Blei**
Columbia University
david.blei@columbia.edu

## A Inference

In this appendix, we describe the details of the approximate inference algorithm for Capsule.

Conditioned on the observed term counts—$n_{dv}$ for vocabulary term $v$ in message $d$; collectively $\mathbf{N}$—our goal is to learn the posterior distribution of the latent variables. Each message is associated with an author entity $a_d$ and a time interval $t_d$ within which that messages was sent. The latent variables are the general topics $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$, the entity topics $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_A$, and the event topics $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_T$, as well as the message-specific strengths $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_D$, $\zeta_1, \ldots, \zeta_D$, and $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_D$, the entity-specific strengths $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_A$ and $\xi_1, \ldots, \xi_A$, and the event strengths $\psi_1, \ldots, \psi_T$. See figures 3 and 4 for the graphical model and generative process.

As for many Bayesian models, the posterior distribution is not tractable to compute; we must instead approximate it. We therefore introduce an approximate inference algorithm for Capsule, based on variational methods (Jordan et al., 1999; Wainwright and Jordan, 2008). Variational methods approximate the true posterior distribution $p$ with a (simpler) variational distribution $q$. Inference then consists of minimizing the KL divergence from $q$ to $p$. This is equivalent to maximizing the evidence lower bound (ELBO):

$$\mathcal{L}(q) = \mathbb{E}_q \left[\log p(\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) - \log q(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi})\right]. \tag{6}$$

We define $q$ using the mean field assumption:

$$\begin{aligned} &q(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) = \\ &\prod_{d=1}^{D} \left( q(\zeta_d \mid \lambda_d) \prod_{k=1}^{K} q(\theta_{dk} \mid \lambda^{\theta}_{dk}) \prod_{t=1}^{T} q(\epsilon_{dt} \mid \lambda^{\epsilon}_{dt}) \right) \times \\ &\prod_{k=1}^{K} \left( q(\boldsymbol{\beta}_k \mid \lambda^{\beta}_k) \prod_{a=1}^{A} q(\phi_{ak} \mid \lambda^{\phi}_{ak}) \right) \prod_{a=1}^{A} \left( q(\boldsymbol{\eta}_a \mid \lambda^{\eta}_a)\, q(\xi_a \mid \lambda^{\xi}_a) \right) \prod_{t=1}^{T} \left( q(\boldsymbol{\gamma}_t \mid \lambda^{\gamma}_t)\, q(\psi_t \mid \lambda^{\gamma}_t) \right) \end{aligned} \tag{7}$$

The variational distributions for the topics $q(\boldsymbol{\beta}_k)$, $q(\boldsymbol{\eta}_a)$, and $q(\boldsymbol{\gamma}_t)$ are all Dirichlet distributions with free variational parameters $\lambda^{\beta}_k$, $\lambda^{\eta}_a$, and $\lambda^{\gamma}_t$, respectively. The variational distributions for the strengths $q(\theta_{dk})$, $q(\zeta_d)$, $q(\epsilon_{dt})$, $q(\phi_{ak})$, $q(\xi_a)$, and $q(\psi_t)$ are all gamma distributions with free variational parameters $\lambda^{\theta}_{dk}$, $\lambda^{\zeta}_d$, $\lambda^{\epsilon}_{dt}$, $\lambda^{\phi}_{ak}$, $\lambda^{\xi}_a$, and $\lambda^{\psi}_t$, respectively. Each of these parameters has two components: shape $s$ and rate $r$.

The expectations under $q$, which we need to maximize the ELBO, have closed analytic forms. We therefore update each free variational parameter in turn, following a standard coordinate-ascent approach.

To obtain update equations for the free variational parameters, we introduce auxiliary latent variables:

$$z_{dkv}^{\mathcal{K}} \sim \text{Poisson}\left(\theta_{dk}\beta_{kv}\right) \tag{8}$$

$$z_{dv}^{\mathcal{A}} \sim \text{Poisson}\left(\zeta_d \eta_{a_d v}\right) \tag{9}$$

$$z_{dtv}^{\mathcal{T}} \sim \text{Poisson}\left(f(t_d, t)\, \epsilon_{dt}\gamma_{tv}\right), \tag{10}$$

where the superscripts $\mathcal{K}$, $\mathcal{A}$, and $\mathcal{T}$ indicate the general, entity, and event topics, respectively. When marginalized out, these variables—collectively $\mathbf{z}$—leave the model intact. Because the Poisson distribution has an additive property, the value of $n_{dv}$ is completely determined by the values of these variables:

$$n_{dv} = \sum_{k=1}^{K} z_{dkv}^{\mathcal{K}} + z_{dv}^{\mathcal{A}} + \sum_{t=1}^{T} z_{dtv}^{\mathcal{T}}. \tag{11}$$

Coordinate-ascent variational inference depends on the conditional distribution of each latent variable given the values of the other latent variables and the data. We use $D(a)$ to denote the set of messages sent by entity $a$ and $D(t)$ to denote the set of messages potentially affected by event $t$ (e.g., all messages sent after time interval $t$, in the case of an exponential decay function). The conditional distributions are:

$$(\boldsymbol{\beta}_k \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Dirichlet}_V\left(\alpha + \sum_{d=1}^{D} z_{dk1}^{\mathcal{K}}, \ldots, \alpha + \sum_{d=1}^{D} z_{dkV}^{\mathcal{K}}\right) \tag{12}$$

$$(\boldsymbol{\eta}_a \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Dirichlet}_V\left(\alpha + \sum_{d \in D(a)} z_{d1}^{\mathcal{A}}, \ldots, \alpha + \sum_{d \in D(a)} z_{dV}^{\mathcal{A}}\right) \tag{13}$$

$$(\boldsymbol{\gamma}_t \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Dirichlet}_V\left(\alpha + \sum_{d \in D(t)} z_{d1t}^{\mathcal{T}}, \ldots, \alpha + \sum_{d \in D(t)} z_{dVt}^{\mathcal{T}}\right) \tag{14}$$

$$(\theta_{dk} \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Gamma}\left(s + \sum_{v=1}^{V} z_{dkv}^{\mathcal{K}},\ \phi_{a_d k} + \sum_{v=1}^{V} \beta_{kv}\right) \tag{15}$$

$$(\zeta_d \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Gamma}\left(s + \sum_{v=1}^{V} z_{dv}^{\mathcal{A}},\ \xi_{a_d} + \sum_{v=1}^{V} \eta_{a_d v}\right) \tag{16}$$

$$(\epsilon_{dt} \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Gamma}\left(s + \sum_{v=1}^{V} z_{dtv}^{\mathcal{T}},\ \psi_t + f(t_d, t) \sum_{v=1}^{V} \gamma_{tv}\right) \tag{17}$$

$$(\phi_{ak} \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Gamma}\left(s + |D(a)|\, s,\ r + \sum_{d \in D(a)} \theta_{dk}\right) \tag{18}$$

$$(\xi_a \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\psi}) \sim \text{Gamma}\left(s + |D(a)|\, s,\ r + \sum_{d \in D(a)} \zeta_d\right) \tag{19}$$

$$(\psi_t \mid \mathbf{N}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}) \sim \text{Gamma}\left(s + |D(t)|\, s,\ r + \sum_{d \in D(t)} \epsilon_{dt}\right). \tag{20}$$

The conditional distribution of the auxiliary latent variables is:

$$(\langle \mathbf{z}^{\mathcal{K}}_{dv}, \mathbf{z}^{\mathcal{A}}_{dv}, \mathbf{z}^{\mathcal{T}}_{dv} \rangle \mid \mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\psi}) \sim \text{Mult}\left(n_{dv}, \boldsymbol{\omega}_{dv}\right), \tag{21}$$

where

$$\boldsymbol{\omega}_{dv} \propto \langle \theta_{d1}\beta_{1v}, \ldots, \theta_{dK}\beta_{Kv}, \zeta_d \eta_{a_d v}, f(t_d, 1)\,\epsilon_{d1}\gamma_{1v}, \ldots, f(t_d, T)\,\epsilon_{dT}\gamma_{Tv}\rangle. \tag{22}$$

Given the conditional distributions, coordinate-ascent variational inference involves setting each free variational parameter to the expected value of the corresponding model parameter under the variational distribution. We provide pseudeocode in algorithm 1; we use $\boldsymbol{\lambda}$ to denote the entire set of free variational parameters and $V(d)$ to denote the set of vocabulary terms present in document $d$. Our approximate inference algorithm produces a fitted variational posterior distribution which can then be used as a proxy for the true posterior distribution. The source code is available online at `https://github.com/ajbc/capsule`.

## B Additional Results

In this appendix, we present non-crucial experimental results for Capsule.

Table 6 lists top documents for an event described in Section 5. Table 7 shows a selection of general topics and Table 8 shows a selection of entity topics.

**Model Sensitivity.** We assessed the sensitivity of our model to three different decay functions $f$: exponential, linear, and step functions. We simulated data for each function and then fit Capsule using every permutation of $f$ and multiple settings for event decay duration. We considered a step function,

$$f(t_d, t) = \begin{cases} 1, & \text{if } t \leq t_d < t + \tau \\ 0, & \text{otherwise,} \end{cases} \tag{23}$$

as well as linear decay,

$$f(t_d, t) = \begin{cases} 1 - \frac{t_d - t}{\tau}, & \text{if } t \leq t_d < t + \tau \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

and an exponential decay function:

$$f(t_d, t) = \begin{cases} 0, & \text{if } t \leq t_d < t + \tau \\ \exp\left\{\frac{-(t_d - t)}{\tau/5}\right\}, & \text{otherwise.}^{1} \end{cases} \tag{25}$$

We used duration $\tau = 3$ and simulated ten data sets for each of the three functions $f$. In fitting the models, we also considered all three functions $f$ and varied the decay duration $\tau$ from 1 to 5. Figure 6 shows the results of these experiments, using both event detection and document recovery metrics discussed previously.

As expected, the model performs best when the model decay function matches the function used to generate the data. For both event detection and document recovery, the exponential decay was least sensitive to the setting of duration $\tau$ used in fitting the data; it was also the least sensitive to the function used in simulating the data. In exploring results on the real-world cable data, we found that the exponential decay provided the most interpretable results.

[1] Unlike the linear and step functions, the exponential function could be evaluated for any time interval $t$ after a document's appearance at $t_d$; the function is truncated for computational reasons. The mean lifetime of this exponential decay is the duration $\tau$ is divided by 5—this ensures that 99.3% of the area under the curve is reached before the function is truncated at duration $\tau$.

**Algorithm 1:** Coordinate-ascent variational inference for Capsule.

**Input**: observed term counts $\mathbf{N}$
**Output**: approximate posterior distribution of the latent variables, in terms of free variational parameters $\boldsymbol{\lambda}$
**Initialize** $\mathbb{E}[\boldsymbol{\beta}_k]$ to slightly random around uniform for each general topic $k$
**Initialize** $\mathbb{E}[$ all other latent variables $]$ to uniform
**for** *iteration* $m = 1, \ldots, M$ **do**
  **set** $\lambda^{\theta,r}$, $\lambda^{\zeta,r}$, and $\lambda^{\epsilon,r}$ to 0 and set remaining $\boldsymbol{\lambda}$ using priors
  **update** $\lambda^{\theta,r}_{dk} \mathrel{+}= \sum_V \mathbb{E}[\boldsymbol{\beta}_{kv}]$ for each message $d$ and general topic $k$
  **for** *message* $d = 1, \ldots, D$ **do**
    **for** *term* $v \in V(d)$ **do**
      **set** $\boldsymbol{\omega}_{dv}$ using expected values of the latent variables (equation (22))
      **set** $\mathbb{E}[\langle \mathbf{z}^{\mathcal{K}}_{dv}, \mathbf{z}^{\mathcal{A}}_{dv}, \mathbf{z}^{\mathcal{T}}_{dv} \rangle] = n_{dv}\boldsymbol{\omega}_{dv}$
      **update** $\lambda^{\beta}_{kv} \mathrel{+}= \mathbb{E}[z^{\mathcal{K}}_{dkv}]$ for each general topic $k$ (equation (12))
      **update** $\lambda^{\eta}_{a_d v} \mathrel{+}= \mathbb{E}[z^{\mathcal{A}}_{dv}]$ (equation (13))
      **update** $\lambda^{\gamma}_{tv} \mathrel{+}= \mathbb{E}[z^{\mathcal{T}}_{dtv}]$ for each time interval $t$ (equation (14))
      **update** $\lambda^{\theta,s}_{dk} \mathrel{+}= \mathbb{E}[z^{\mathcal{K}}_{dkv}]$ for each general topic $k$ (equation (15))
      **update** $\lambda^{\zeta,s}_{d} \mathrel{+}= \mathbb{E}[z^{\mathcal{A}}_{dv}]$ (equation (16))
      **update** $\lambda^{\epsilon,s}_{dt} \mathrel{+}= \mathbb{E}[z^{\mathcal{K}}_{dtv}]$ for each time interval $t$ (equation (17))
    **end**
    **set** $\lambda^{\theta,r}_{dk} = \mathbb{E}[\phi_{a_d k}] + \sum_v \mathbb{E}[\beta_{kv}]$ for each general topic $k$ (equation (15))
    **set** $\lambda^{\zeta,r}_{d} = \mathbb{E}[\xi_{a_d}] + \sum_v \mathbb{E}[\eta_{a_d v}]$ (equation (16))
    **set** $\lambda^{\epsilon,r}_{dt} = \mathbb{E}[\psi_t] + f(t_d, t) \sum_v \mathbb{E}[\gamma_{tv}]$ for each time interval $t$ (equation (17))
    **set** $\mathbb{E}[\theta_{dk}] = \lambda^{\theta,s}_{dk} \,/\, \lambda^{\theta,r}_{dk}$ for each general topic $k$
    **set** $\mathbb{E}[\zeta_d] = \lambda^{\zeta,s}_{d} \,/\, \lambda^{\zeta,r}_{d}$
    **set** $\mathbb{E}[\epsilon_{dt}] = \lambda^{\epsilon,s}_{dt} \,/\, \lambda^{\epsilon,r}_{dt}$ for each time interval $t$
    **update** $\lambda^{\phi,s}_{a_d k} \mathrel{+}= s$ for each general topic $k$ (equation (18))
    **update** $\lambda^{\xi,s}_{a_d} \mathrel{+}= s$ (equation (19))
    **update** $\lambda^{\psi,s}_{t} \mathrel{+}= s$ for each time interval $t$ where $f(t_d, t) \neq 0$ (equation (20))
    **update** $\lambda^{\phi,r}_{a_d k} \mathrel{+}= \theta_{dk}$ for each general topic $k$ (equation (18))
    **update** $\lambda^{\xi,r}_{a_d} \mathrel{+}= \zeta_d$ (equation (19))
    **update** $\lambda^{\psi,r}_{t} \mathrel{+}= \epsilon_{dt}$ for each time interval $t$ (equation (20))
  **end**
  **set** $\mathbb{E}[\boldsymbol{\beta}_k] = \boldsymbol{\lambda}^{\beta}_{k} \,/\, \sum_v \lambda^{\beta}_{kv}$ for each general topic $k$
  **set** $\mathbb{E}[\boldsymbol{\eta}_a] = \boldsymbol{\lambda}^{\eta}_{a} \,/\, \sum_v \lambda^{\eta}_{av}$ for each entity $a$
  **set** $\mathbb{E}[\boldsymbol{\gamma}_t] = \boldsymbol{\lambda}^{\gamma}_{t} \,/\, \sum_v \lambda^{\gamma}_{tv}$ for each time interval $t$
  **set** $\mathbb{E}[\phi_{ak}] = \lambda^{\phi,s}_{ak} \,/\, \lambda^{\phi,r}_{ak}$ for each entity $a$ and general topic $k$
  **set** $\mathbb{E}[\xi_a] = \lambda^{\xi,s}_{a} \,/\, \lambda^{\xi,r}_{a}$ for each entity $a$
  **set** $\mathbb{E}[\psi_t] = \lambda^{\psi,s}_{t} \,/\, \lambda^{\psi,r}_{t}$ for each time interval $t$
**end**
**return** $\boldsymbol{\lambda}$

| $f * \epsilon$ | **Date** | **Entity** | **Subject** |
|---|---|---|---|
| 6.86 | 1976-07-07 | Cairo | Possible SC meeting on Israeli rescue operation |
| 6.18 | 1976-07-10 | Kuwait | Media reaction to Bicentennial summary |
| 6.15 | 1976-07-06 | Damascus | Syria condemns Israeli operation to free Air France ... |
| 5.91 | 1976-07-08 | Tel Aviv | Passengers comment on Air France hijacking |
| 5.89 | 1976-07-06 | Stockholm | Possible SC meeting on Israeli rescue operation |
| 5.38 | 1976-07-09 | Nicosia | Bicentennial activities in Cyprus |
| 5.09 | 1976-07-11 | State | Security Council debate on Entebbe events CONFID... |
| 4.77 | 1976-07-09 | State | Travel of Peter M. Storm, House Budget Committee |
| 4.76 | 1976-07-06 | Jidda | Weekly Saudi Editorial Summary (June 30-July 6) |
| 4.68 | 1976-07-08 | Lusaka | SWAPO President seeks assessment of Kissinger-Vor... |
| 4.56 | 1976-07-07 | Stockholm | Ugandan role in Air France hijacking |
| 4.45 | 1976-07-06 | Karachi | Transitional quarter funding for RSS travel |
| 4.43 | 1976-07-06 | Athens | Bicentennial anniversary in Greece |
| 4.37 | 1976-07-08 | Damascus | Beirut travel |
| 4.34 | 1976-07-10 | State | Status of Mrs. Bloch |
| 4.17 | 1976-07-07 | Hong Kong | Hong Kong Communist press denounces Israeli resc... |
| 4.12 | 1976-07-08 | Dar es Salaam | President Nyerere's fourth of July messages |
| 4.09 | 1976-07-10 | Moscow | Pravda and Krasnaya Zvezda on Entebbe rescue oper... |

**Table 6:** Top documents for the week after the US bicentennial celebration and Operation Entebbe. Capsule identifies documents relevant to both these real-world events.

## References


Zoubin Ghahramani and Matthew J Beal. 2001. Propagation algorithms for variational bayesian learning. *Advances in Neural Information Processing Systems (NIPS)*, pages 507–513.

Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. 1999. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, November.

Martin J. Wainwright and Michael I. Jordan. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, January.

| top terms |
|---|
| church, vatican, catholic, bishop, pope, ford, cardinal, ban, religious, archbishop |
| program, university, grant, education, school, post, institute, research, center, american |
| security, council, terrorist, threat, sc, sabotage, protective, herein, unsc, honour |
| visit, hotel, schedule, arrival, arrive, depart, please, meet, day, room |
| labor, union, strike, ilo, employment, federation, afl cio, trade, worker, confederation |
| bank, credit, loan, investment, finance, payment, financial, eximbank, opic, central |
| law, case, court, legal, investigation, arrest, justice, sentence, trial, attorney |
| party, government, election, opposition, national, leader, campaign, vote, support, anti |
| tax, company, pay, lease, compensation, exemption, repatriation, income, taxation, fee |
| oil, petroleum, opec, crude, gulf, price, exploration, refinery, energy, company |
| israel, arab, israeli, middle, egypt, peace, plo, cairo, egyptian, lebanon |
| radio, television, broadcast, allotment, appropriation, obligation, zero, warc, transmitter, network |
| india, indian, pakistan, delhi, goi, ocean, bangladesh, transit, pakistani, afghan |
| turkish, turkey, cyprus, greek, greece, athens, ankara, morocco, cypriot, algeria |
| aid, relief, emergency, usaid, disaster, donor, wfp, sahel, ifad, unicef |
| aircraft, team, flight, clearance, transport, civair, aviation, traffic, charter, cargo |
| soviet, moscow, press, ussr, soviet union, american, one, war, communist, article |
| sea, zone, marine, maritime, fish, coastal, continental, territorial, mile, fishery |

**Table 7:** Top vocabulary terms for a selection of general topics, one per row, according to topic distributions $\beta_k$. Capsule identifies general diplomatic themes that can be relevant to any entity.

| entity | top terms |
|---|---|
| Ankara | turkish, turkey, ankara, government, cyprus, greek, party, one, time |
| Athens | greek, athens, greece, gog, government, cyprus, turkish, press, minister |
| Auckland | new zealand, company, box, trade, contact, opportunity, united states |
| Baghdad | iraqi, iraq, goi, arab, state, regime, ministry, government, party |
| Berlin | berlin, frg, german, senat, time, bonn, trade, one, agreement |
| Bern | swiss, bern, federal, bank, snb, gold, end, interest, national |
| Brussels | belgian, belgium, brussels, government, firestone, european, ministry |
| Budapest | hungarian, hungary, trade, mudd, one, time, puja, well, policy |
| Buenos Aires | argentine, argentina, goa, us, hill, government, one, press, police |
| Cairo | egyptian, cairo, egypt, arab, israeli, israel, peace, agreement, president |
| Canberra | australian, australia, goa, government, minister, whitlam, end, dfa, time |
| Dakar | senegalese, president, african, summary, conference, end, support, one |
| Dar es Salaam | tangov, salaam, tanzanian, spain, president, government, african, one |
| Guayaquil | ecuador, ecuadorean, port, congen, one, tuna, local, time, boat |
| Islamabad | pakistan, gop, government, one, party, minister, general, opposition, ppp |
| Paris | paris, france, rush, french, one, government, amconsul, quai, european |
| Jerusalem | jerusalem, bank, israeli, us, israel, plo, one, arab, unifil |
| Jidda | saudi, jidda, saudi arabia, prince, us, fahd, one, time, government |
| Johannesburg | black, africa, african, trade, union, police, labor, one, committee |
| Kabul | afghan, government, goa, minister, one, pakistan, regime, time, ministry |
| Lima | peru, gop, lima, peruvian, dean, minister, general, marcona, government |
| Lisbon | portugal, portuguese, gop, lisbon, government, party, summary, minister |
| London | london, british, government, fco, labor, agreement, one, washdc, summary |
| Madrid | spanish, spain, madrid, one, govt, general, committee, government, time |
| Nairobi | kenya, nairobi, marshall, embassy, kenyan, unep, le, ref, state |
| Oslo | norwegian, norway, soviet, government, minister, ministry, policy |
| Ottawa | canadian, canada, goc, ottawa, us, extaff, government, minister, federal |
| Peking | chinese, peking, uslo, china, people, teng, one, trade, delegation, hong |
| Phnom penh | penh, phnom, khmer, rice, fank, enemy, cambodia, government, dean |
| Prague | czechoslovak, goc, czech, trade, embassy, one, mfa, time, cssr |
| Quito | ecuador, ecuadorean, gulf, government, minister, bloomfield, general, one |
| Sao Paulo | paulo, brazil, state, brazilian, president, government, congen, one, do |
| Seoul | korea, korean, rok, rokg, seoul, park, government, president, time |
| Singapore | singapore, asean, minister, government, one, prime, comment, vietnam |
| Sofia | bulgarian, trade, one, agreement, american, visit, committee, party |
| Sydney | australia, australian, one, general, american, state, government, post |
| Tokyo | japan, japanese, tokyo, fonoff, summary, miti, end, diet, time |
| Taipei | taiwan, groc, china, chinese, government, american, one, local, republic |
| The Hague | dutch, netherlands, hague, government, minister, party, stoel, mfa, one |
| USUN New York | committee, usun, priority, report, draft, resolution, sc, comite, rep, new york |
| Vancouver | canada, government, canadian, british, columbia, pipeline, federal, editorial |
| Zagreb | yugoslav, yugoslavia, croatian, fair, belgrade, american, one, ina, summary |
| Zurich | swiss, congen, consulate, general, american, bern, dollar, shipment |

**Table 8:** Top vocabulary terms for a selection of entities according to entity-exclusive topics $\eta_n$. Capsule identifies entity-specific themes and interests.

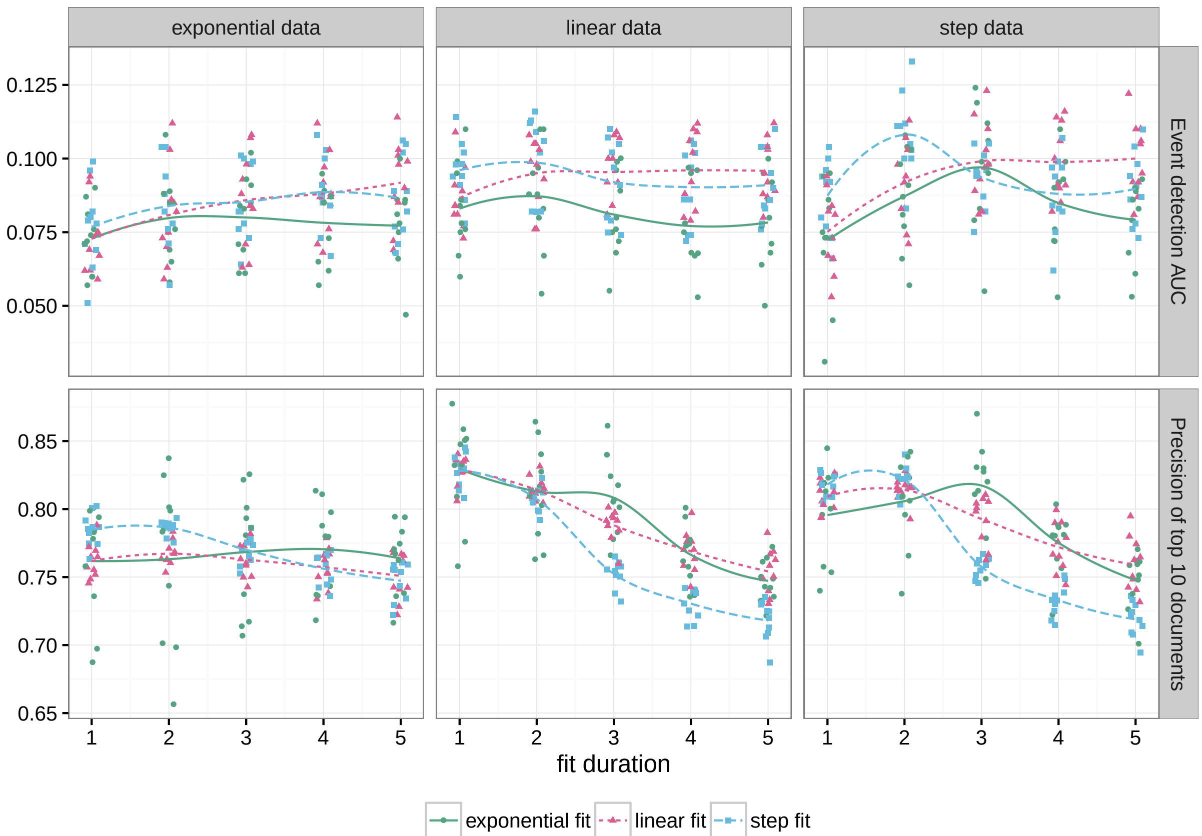


**Figure 6:** Assessment of model parameter sensitivity on simulated data–Capsule performs best when the model decay function matches the function used to generate the data. The exponential decay is least sensitive to the setting of duration $\tau$ and the true function $f$.