

Development and Derivation of the Psychological Model

July 23, 2018

Dramatic real-world events are known to have the power to impact on public opinions and to cause shift on public attitudes. For example, assassinations of Martin Luther King created a cultural shift in attitudes on race issues. Thus we are motivated to analyze news and user generated contents on social networking platforms, model the evolution and shifts of public opinions, extract common patterns and explain anomalies.

This document provides development logs of the evolutionary summarization model, including intuitions, derivations, and result summaries.

Part I

Opinion Models

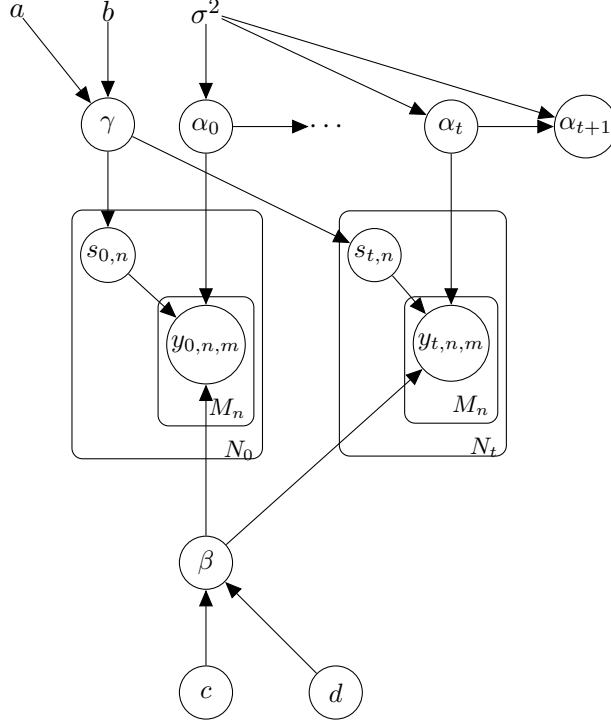


Figure 1: Plate notation of the proposed opinion evolution model

We first explore our options on modeling the evolution of public opinions. Before we go into the details, we have to list a few assumptions here. (1) We only consider bi-polarized opinions, i.e. each unit piece of social comment is pre-processed by some opinion classifiers to be either positive or negative. The unit to be processed can be a phrase about an entity, a sentence, or a minimal length semantic unit. (2) We consider there is a background opinion distribution, i.e. how users normally react to a certain entity or a particular event. The background is smoothly and slowly changing, e.g. the public opinion for civil rights is constantly changing. (3) However, sometimes a new piece of evidence might trigger a sudden shift on public opinions, e.g. the assassination of Martin Luther King boosts public supports to civil rights.

First we generate the prior distribution for the background opinion distributions.

- For time $t = 0$, sample for the public opinion distribution, $\alpha_0 \sim \mathcal{N}(0, \sigma^2 I)$.

- For items $t = 1 : N$, sample $\alpha_{t+1} \sim \mathcal{N}(\alpha_t, \sigma^2)$. As \mathcal{N} is a continuous and differentiable distribution, the evolution of background opinions is smooth and slow.
- Generate a global prior for the switch, i.e. a variable that controls how likely the public opinion to change, by $\gamma \sim \text{Beta}(a, b)$

Note that all α s are vectors of real numbers, $\alpha \in \mathcal{R}^2$. Next to generate the opinion distribution, we adopt $\pi(\cdot)$ function to convert α to a value within $(0, 1)$, $\pi(\alpha_t) = \frac{\exp \alpha_{t,0}}{\exp \alpha_{t,0} + \exp \alpha_{t,1}}$.

For time t , with N_t pieces of evidence, i.e. N_t piece of news report published at time t , each evidence n contains M_n observations, e.g. user comments, we mimic the following generation process.

- For each pieces of evidence
 - Generate a switch $s_t \sim \text{Bern}(\gamma)$
 - For each observation, generate $y_{t,n,m} \sim \begin{cases} \text{Bern}(\pi(\alpha_t)) & \text{if } s_{t,n} = 1 \\ \text{Bern}(\beta) & \text{if } s_{t,n} = 0 \end{cases}$

The joint probability is given by

$$p(\gamma, \beta, \alpha_0, \dots, \alpha_T, \vec{s}, \vec{y}, |a, b, c, d, \sigma^2) \\ = p(\gamma|a, b)p(\beta|c, d)p(\alpha_{0:T}|\sigma^2) \prod_t \prod_n p(s_{t,n}|\gamma) \prod_m p(y_{t,n,m}|s_{t,n}, \alpha_t, \beta)$$

However, as the above complete likelihood involves natural parameters $\pi(\alpha)$, we apply the lower-bound $\forall t, \ln(\exp \alpha_{t,0} + \exp \alpha_{t,1}) \leq \ln \hat{\xi}_t + \frac{\exp \alpha_{t,0} + \exp \alpha_{t,1} - \hat{\xi}_t}{\hat{\xi}_t}$ to $p(y_{t,n,m}|s_{t,n}, \alpha_t, \beta) = [\pi(\alpha_t)^{y_{t,n,m}} (1 - \pi(\alpha_t))^{1 - y_{t,n,m}}]^{s_{t,n}} [\beta^{y_{t,n,m}} (1 - \beta)^{1 - y_{t,n,m}}]^{1 - s_{t,n}}$ whenever necessary. In the nutshell, the optimization algorithm is variational inference. Thus we make the following assumptions.

$$q(Z|\vec{y}, a, b, c, d, \sigma^2) = q(\gamma|\hat{a}, \hat{b})q(\beta|\hat{c}, \hat{d})q(\alpha_{0:T}|\alpha_{\hat{0}:T}) \prod_{t,n} q(s_{t,n}|e_{t,n}),$$

where Z includes all hidden variables, $e_{t,n} \in \mathcal{R}^2$ is a vector.

We implement the iterations over all hidden variables.

$$\begin{aligned} \ln q(\gamma|\hat{a}, \hat{b}) &= \mathbb{E}[\ln p(\gamma|a, b) + \sum_{t,n} \ln p(s_{t,n}|\gamma)] + \text{const} \\ &= (a - 1) \ln \gamma + (b - 1) \ln(1 - \gamma) + \sum_{t,n} \mathbb{E}[s_{t,n}] \ln \gamma + \sum_{t,n} \mathbb{E}[1 - s_{t,n}] \ln(1 - \gamma) + \text{const} \end{aligned}$$

$$\begin{aligned}
\hat{a} &= a + \sum_{t,n} \mathbb{E}[s_{t,n}] \\
\hat{b} &= b + \sum_{t,n} \mathbb{E}[1 - s_{t,n}] \\
\mathbb{E}[\gamma] &= \frac{\hat{a}}{\hat{a} + \hat{b}} \\
\mathbb{E}[\ln \gamma] &= \phi(\hat{a}) - \phi(\hat{a} + \hat{b}) \\
\mathbb{E}[\ln(1 - \gamma)] &= \phi(\hat{b}) - \phi(\hat{a} + \hat{b})
\end{aligned} \tag{1}$$

Note that the last conclusion is obtained by using $\lim_{k \leftarrow \infty} \frac{\delta(1-x)^k}{\delta k} = \ln(1-x)$.

$$\begin{aligned}
\ln q(\beta|\hat{c}, \hat{d}) &= \mathbb{E}[\ln p(\beta|c, d) + \sum_{t,n,m} \ln p(y_{t,n,m}|s_{t,n}, \alpha_t, \beta)] + \text{const} \\
&= (c-1) \ln \beta + (d-1) \ln(1-\beta) + \sum_{t,n,m} \mathbb{E}[1 - s_{t,n}] (y_{t,n,m} \ln \beta + (1 - y_{t,n,m}) \ln(1-\beta)) + \text{const}
\end{aligned}$$

$$\begin{aligned}
\hat{c} &= c + \mathbb{E}[1 - s_{t,n}] \sum_{t,n,m} y_{t,n,m} \\
\hat{d} &= d + \mathbb{E}[1 - s_{t,n}] \sum_{t,n,m} (1 - y_{t,n,m}) \\
\mathbb{E}[\beta] &= \frac{\hat{c}}{\hat{c} + \hat{d}} \\
\mathbb{E}[\ln \beta] &= \phi(\hat{c}) - \phi(\hat{c} + \hat{d}) \\
\mathbb{E}[\ln(1 - \beta)] &= \phi(\hat{d}) - \phi(\hat{c} + \hat{d})
\end{aligned} \tag{2}$$

$$\begin{aligned}
\ln q(s_{t,n}|e_{t,n}) &= \mathbb{E}[\ln p(s_{t,n}|\gamma) + \sum_m \ln p(y_{t,n,m}|s_{t,n}, \alpha_t, \beta)] + \text{const} \\
&= s_{t,n} \mathbb{E}[\ln \gamma] + (1 - s_{t,n}) \mathbb{E}[\ln(1 - \gamma)] + s_{t,n} \sum_m (y_{t,n,m} \mathbb{E}[\ln \pi(\alpha_t)] + (1 - y_{t,n,m}) \mathbb{E}[\ln \pi(\alpha_t)]) \\
&\quad + (1 - s_{t,n}) \sum_m (y_{t,n,m} \mathbb{E}[\ln \beta] + (1 - y_{t,n,m}) \mathbb{E}[\ln(1 - \beta)]) + \text{const}
\end{aligned}$$

$$\begin{aligned}
\ln e_{t,n_0} &= \phi(\hat{a}) - \phi(\hat{a} + \hat{b}) + \sum_m y_{t,n,m} \mathbb{E}[\alpha_{t,0}] + \sum_m (1 - y_{t,n,m}) \mathbb{E}[\alpha_{t,1}] \\
&= \phi(\hat{a}) - \phi(\hat{a} + \hat{b}) + \sum_m y_{t,n,m} \alpha_{t,0} + \sum_m (1 - y_{t,n,m}) \alpha_{t,1} \\
\ln e_{t,n_1} &= \phi(\hat{b}) - \phi(\hat{a} + \hat{b}) + \sum_m y_{t,n,m} (\phi(\hat{c}) - \phi(\hat{c} + \hat{d})) + \sum_m (1 - y_{t,n,m}) (\phi(\hat{d}) - \phi(\hat{c} + \hat{d}))
\end{aligned} \tag{3}$$

We have to renormalize e_{t,n_0} , e_{t,n_1} to make sure $e_{t,n_0} + e_{t,n_1} = 1$.

Finally, for α , we extract the ELBO involving α terms, we adopt variational Kalman Filtering. The state space model is $q(\alpha|_t \hat{\alpha}_t) \sim \mathcal{N}(\hat{\alpha}_t, \sigma^2 I)$. We have the following results.

$$m_t H(q) = \frac{1}{2} [1 + \ln(2\sigma^2\pi)]$$

$$\begin{aligned}
L(\alpha_{0:T}) &= \mathbb{E}[\ln p(\alpha_{0:T} | \alpha_{0:T}, \sigma^2) + \sum_{t,n,m} \ln p(y_{t,n,m} | s_{t,n}, \alpha_t, \beta)] + H(q(\alpha)) + \text{const} \\
L(\alpha_t) &= -\frac{1}{2\sigma^2} \|\tilde{m}_t - \tilde{m}_{t-1}\|^2 - \frac{1}{\sigma^2} \text{Tr}(\tilde{V}_t) + \frac{1}{2\sigma^2} (\text{Tr}(\tilde{V}_0) - \text{Tr}(\tilde{V}_T)) \\
&\quad \sum_{t,n,m} \mathbb{E}[s_{t,n} y_{t,n,m} \ln \pi(\alpha_{t,0}) + s_{t,n} (1 - y_{t,n,m}) \ln(\pi(\alpha_{t,1}))] \\
&\quad + \frac{1}{2} \sum_t [\ln |\tilde{V}_t| + T \ln 2\pi] + \text{const} \\
&\geq -\frac{1}{2\sigma^2} \|\tilde{m}_t - \tilde{m}_{t-1}\|^2 - \frac{1}{\sigma^2} \text{Tr}(\tilde{T}_t) + \frac{1}{2\sigma^2} (\text{Tr}(\tilde{V}_0) - \text{Tr}(\tilde{V}_T)) \\
&\quad + \sum_t (\sum_n \mathbb{E}[s_{t,n}] \sum_m y_{t,n,m} \alpha_{t,0} + \sum_t (\sum_n \mathbb{E}[s_{t,n}] \sum_m (1 - y_{t,n,m})) \alpha_{t,1} \\
&\quad - \sum_{t,n} \mathbb{E}[s_{t,n}] \ln \hat{\xi}_t - \sum_{t,n} \mathbb{E}[s_{t,n}] \frac{\mathbb{E}[\exp \alpha_{t,0} + \exp \alpha_{t,1}]}{\hat{\xi}_t} + \text{const}
\end{aligned}$$

Let $\frac{\delta(L(\alpha_t))}{\delta \alpha_t} = 0$, we have should use the forward-backward recursive algorithm in variational Kalman filtering.

The ELBO involving $\hat{\xi}_t$ terms are

$$L(\hat{\xi}_t) = -\sum_{t,n} \mathbb{E}[s_{t,n}] \ln \hat{\xi}_t - \sum_{t,n} \mathbb{E}[s_{t,n}] \frac{\mathbb{E}[\exp \alpha_{t,0} + \exp \alpha_{t,1}]}{\hat{\xi}_t}$$

As in the variational Kalman filtering $\mathbb{E}[\exp \alpha_{t,0} + \exp \alpha_{t,1}] = \sum_{i=1,2} \exp(\tilde{m}_{t,i} + \tilde{V}_{t,i}/2)$, we have:

$$\begin{aligned}
\frac{\delta - L(\hat{\xi}_t)}{\delta \xi_t} &= \frac{\sum_{t,n} \mathbb{E}[s_{t,n}]}{\hat{\xi}_t} + \frac{\sum_{t,n} \mathbb{E}[s_{t,n}] \sum_{i=1,2} \exp(\tilde{m}_{t,i} + \tilde{V}_{t,i}/2)}{\hat{\xi}_t^2} \quad (4) \\
&= 0 \\
\hat{\xi}_t &= \sum_{i=1,2} \exp(\tilde{m}_{t,i} + \tilde{V}_{t,i}/2)
\end{aligned}$$