IPTM Issues

Bomin Kim

January 24, 2018

1 Interaction-pattern specific statsitics?

The draft says we use $\boldsymbol{x}_{adrc}$ for recipient generating process, and $\boldsymbol{y}_{adc}$ cor timestmaps generating process. However, we currently do not use $\boldsymbol{y}_{adc}$. Instead, we use $\boldsymbol{y}_{ad}$—sender-specific intercepts and two time indicators (weekends and AM/PM).

Q1. Is it fine to use non-IP-specific statistis? No identifiability issues?

Answer: ?

newline **Q2.** If it is fine, can we switch to $\boldsymbol{x}_{adr}$—where we maintain all definition of network statistics but drop c by not accounting for the proportions—as well?

Reason: Since we use $\boldsymbol{x}_{adrc}$ via $\frac{N_{dc}}{N_d}$, every topic-token assignment (i.e. update of z_{dn}) and interaction-pattern assignment require re-calculation of "entire" history statistics. Here is how it works in topic-token assignment:

for (d in 1:D) {
for (w in 1:N_d) {
for (k in 1:K) {
Assume $z_{dn} = k$ then (if this changes N_{dc}) compute $\frac{N_{dc}}{N_d}$ again
Re-calculate x_{adrc} for $d = d^*$ where d^* is the last document affected by d ($t_{d^*} \approx t_d$+384)
Re-calculate λ_{adr} for $d = d^*$ where d^* is the last document affected by d ($t_{d^*} \approx t_d$+384)
Re-calculate μ_{adc} for d
$P(z_{dn} = k) \propto$ Equation (15) evaluated for d^*
}
}
}

However, if we drop c in x_{adrc}, we do not need to re-calculate the network statistics—they are truely observed given the corpus. Only λ_{adr} and μ_{ad} will vary by topic-token assignments and interaction-pattern assignments, and these are not computationally heavy. Currently, one outer iteration (given 5 inner updates for M-H) takes around 400 secs in my machine, mostly due to the network statistics updates.

2 PPE— from generative proces or full conditional distributions?

I thought we arrived at the conclusion to use "full conditional distributions" for posterior predictive experiments. However, due to our definition of network statistics:

$$
\begin{aligned}
&P(a_d = a|\boldsymbol{r}_d, t_d, \boldsymbol{w}_d, \boldsymbol{l}, \boldsymbol{z}, \boldsymbol{b}, \boldsymbol{\eta}, \delta)\\
&\propto \text{(how likely the sender is "a" given the recipeintsfor document d)}\\
&\times \text{(how likely the sender is "a" given the timestamps for document d)}\\
&\times \text{(how this assignment changes the future likelihoods for document } d+1, ..., d^*)\\
&\propto \frac{\exp\Big\{\sum\limits_{r\neq a}(\delta+\lambda_{adr})r_{adr}\Big\}}{Z(\delta, \boldsymbol{\lambda}_{ad})} \times \varphi_\tau(\tau_d; \mu_{ad}, \sigma_\tau^2)\prod_{a'\neq a}\big(1-\Phi_\tau(\tau_d; \mu_{a'd}, \sigma_\tau^2)\big)\\
&\times \prod_{d=d+1}^{d^*}\Big(\prod_{a=1}^{A}\frac{\exp\Big\{\log\big(\mathrm{I}(\|\boldsymbol{u}_{ad}\|_1>0)\big)+\sum\limits_{r\neq a}(\delta+\lambda_{adr})u_{adr}\Big\}}{Z(\delta, \boldsymbol{\lambda}_{ad})}\\
&\quad \times \varphi_\tau(\tau_d; \mu_{a_d d}, \sigma_\tau^2)\prod_{a\neq a_d}\big(1-\Phi_\tau(\tau_d; \mu_{ad}, \sigma_\tau^2)\big)\Big)\\
&\text{via new calculation of } x_{adrc} \text{ for } d = d+1, \dots, d^* \text{ assuming } a_d = a,
\end{aligned} \tag{1}
$$

for all $a = 1, ..., A$ and use multinomial sampling. Very complicated but still do-able. Similarly for recipient predictions,

$$
\begin{aligned}
&P(r_{dr} = 1|a_d, t_d, \boldsymbol{w}_d, \boldsymbol{l}, \boldsymbol{z}, \boldsymbol{b}, \boldsymbol{\eta}, \delta)\\
&\propto \text{(how likely the missing recipient element is "1" given the author of document d)}\\
&\times \text{(how this assignment changes the future likelihoods for document } d+1, ..., d^*)\\
&\propto \exp\{\delta+\lambda_{adr}\} \times \prod_{d=d+1}^{d^*}\Big(\prod_{a=1}^{A}\frac{\exp\Big\{\log\big(\mathrm{I}(\|\boldsymbol{u}_{ad}\|_1>0)\big)+\sum\limits_{r\neq a}(\delta+\lambda_{adr})u_{adr}\Big\}}{Z(\delta, \boldsymbol{\lambda}_{ad})}\Big)\\
&\text{via new calculation of } x_{adrc} \text{ for } d = d+1, \dots, d^* \text{ assuming } r_{dr} = 1,
\end{aligned} \tag{2}
$$

and

$$
\begin{aligned}
&P(r_{dr} = 0|a_d, \boldsymbol{r}_d, \boldsymbol{w}_d, \boldsymbol{l}, \boldsymbol{z}, \boldsymbol{b}, \boldsymbol{\eta}, \delta)\\
&\propto \text{(how likely the missing recipient element is "1" given the author of document d)}\\
&\times \text{(how this assignment changes the future likelihoods for document } d+1, ..., d^*)\\
&\propto \mathrm{I}(\|\boldsymbol{u}_{ad\backslash r}\|_1>0) \times \prod_{d=d+1}^{d^*}\Big(\prod_{a=1}^{A}\frac{\exp\Big\{\log\big(\mathrm{I}(\|\boldsymbol{u}_{ad}\|_1>0)\big)+\sum\limits_{r\neq a}(\delta+\lambda_{adr})u_{adr}\Big\}}{Z(\delta, \boldsymbol{\lambda}_{ad})}\Big)\\
&\text{via new calculation of } x_{adrc} \text{ for } d = d+1, \dots, d^* \text{ assuming } r_{dr} = 0,
\end{aligned} \tag{3}
$$

where this time we don't need to account for time-parts since there is no connection between recipients and timestamps. Again, we can use multinomial between (0/1).

However, when it comes to the conditional distribution of timestamps,

$$
\begin{aligned}
&P(\tau_d = \tau | a_d, t_d, \boldsymbol{w}_d, \boldsymbol{l}, \boldsymbol{z}, \boldsymbol{b}, \boldsymbol{\eta}, \delta) \\
&\propto \text{(how likely the missing time-increment is ``}\tau\text{'' given the author of document d)} \\
&\times \text{(how this assignment changes the future likelihoods for document } d+1, ..., d^*) \\
&\propto \Big(\varphi_\tau(\tau_d; \mu_{a_d d}, \sigma_\tau^2) \times \prod_{a \neq a_d} \big(1 - \Phi_\tau(\tau_d; \mu_{ad}, \sigma_\tau^2)\big)\Big) \\
&\times \prod_{d=d+1}^{d^*} \Big(\prod_{a=1}^{A} \frac{\exp\Big\{\log\big(\mathrm{I}(\|\boldsymbol{u}_{ad}\|_1 > 0)\big) + \sum_{r \neq a} (\delta + \lambda_{adr}) u_{adr}\Big\}}{Z(\delta, \boldsymbol{\lambda}_{ad})} \\
&\times \varphi_\tau(\tau_d; \mu_{a_d d}, \sigma_\tau^2) \prod_{a \neq a_d} \big(1 - \Phi_\tau(\tau_d; \mu_{ad}, \sigma_\tau^2)\big)\Big) \\
&\text{via new calculation of } x_{adrc} \text{ and } y_{adc} \text{ for } d = d+1, \ldots, d^* \text{ assuming } \tau_d = \tau.
\end{aligned}
\tag{4}
$$

Since this is not discrete, Any way to directly sampling missing value of τ from this distribution? I misunderstood that this last sampling can be simply reduced to $\tau \sim \text{lognormal}(\mu_{a_d d}, \sigma_\tau^2)$, but it is not. Possibly M-H, Slice, Hamiltonian, or rejection sampling...?

To impute the missing values from generative process (instead of full conditionals), we need to assume that "this document is the last document". This makes things simpler since we do not need to recalculate future covariates and include $\prod_{d=d+1}^{d^*}$ term in the sampling equations. For example, we could go back to the old experiment—predicting sender, recipient, timestamps jointly for d^{pred} given the $d = 1, \ldots, d^{pred} - 1$ entirely observed such that we simply follow the generative process—although extremely time consuming given that we need to run inference on $d = 1, \ldots, d^{pred} - 1$ for every prediction.