

Supplementary for Gamma-Poisson Dynamic Matrix Factorization Embedded with Metadata Influence

Trong Dinh Thac Do
Advanced Analytics Institute
University of Technology Sydney
thacddtd@gmail.com

Longbing Cao
Advanced Analytics Institute
University of Technology Sydney
longbing.cao@gmail.com

Here we present additional details to complement with the paper in terms of the graphical model of mGDMF, the variational update for item's latent variables, the batch inference algorithm, the full derivation of mGDMF.

1 Graphical Model of mGDMF

We show the graphical model of mGDMF in Figure 1

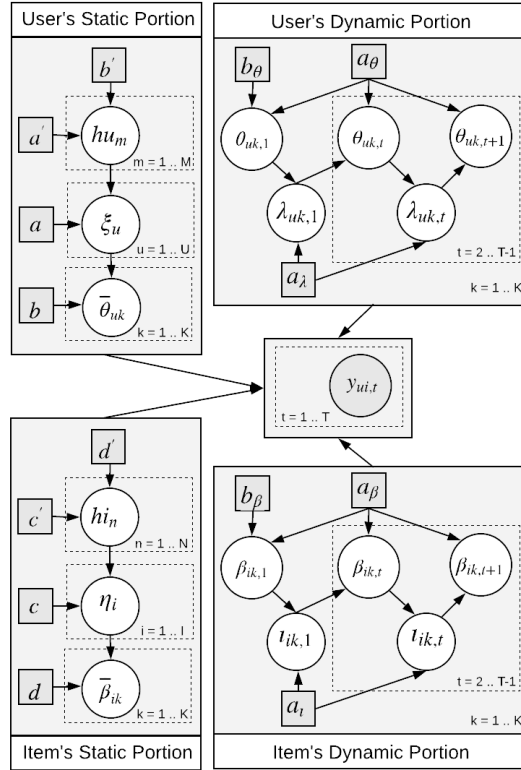


Figure 1: Graphical Model of mGDMF

2 Variational Update for Items Latent Variables

We show the latent variables for items (i.e., hi_n , η_i , $\bar{\beta}_{ik}$, β_{ik} , $\iota_{ik,t}$) with their variational update as mentioned in Table 1 in the paper.

Table 1: Latent Variables, Type, Variational Variables and Variational Update for Items (i.e., hi_n , η_i , $\bar{\beta}_{ik}$, β_{ik} , $\iota_{ik,t}$). χ_n is the number of items having the n^{th} attribute and K is the number of latent components.

Latent Variable	Type	Variational Variable	Variational Update
hi_n	Gamma	ρ_n^{shp} ρ_n^{rte}	$c' + \chi_n c$ $d' + \sum_i \frac{\tau_i^{shp}}{\tau_i^{rte}}$
η_i	Gamma	τ_i^{shp} τ_i^{rte}	$c + Kd$ $\prod_{n=1}^N (\frac{\rho_n^{shp}}{\rho_n^{rte}})^{f_{i,n}} + \sum_k \frac{\bar{\mu}_{ik}^{shp}}{\bar{\mu}_{ik}^{rte}}$
$\bar{\beta}_{ik}$	Gamma	$\bar{\mu}_{ik}^{shp}$ $\bar{\mu}_{ik}^{rte}$	$d + \sum_{u,t} y_{ui,t} \phi_{ui,t,k}$ $\frac{\tau_i^{shp}}{\tau_i^{rte}} + \sum_u (\frac{\bar{\nu}_{uk}^{shp}}{\bar{\nu}_{uk}^{rte}} + \sum_t \frac{\nu_{uk,t}^{shp}}{\nu_{uk,t}^{rte}})$
$\beta_{ik,t}$	Gamma	$\mu_{ik,t}^{shp}$ $\mu_{ik,1}^{rte}$ $\mu_{ik,t,(t>1)}^{rte}$	$a_\beta + a_\iota + \sum_u y_{ui,t} \phi_{ui,t,k}$ $a_\beta b_\beta + a_\iota \frac{\omega_{ik,1}^{shp}}{\omega_{ik,1}^{rte}} + \sum_u (\frac{\bar{\nu}_{uk}^{shp}}{\bar{\nu}_{uk}^{rte}} + \frac{\nu_{uk,1}^{shp}}{\nu_{uk,1}^{rte}})$ $a_\beta \frac{\omega_{ik,t-1}^{shp}}{\omega_{ik,t-1}^{rte}} + a_\iota \frac{\omega_{ik,t}^{shp}}{\omega_{ik,t}^{rte}} + \sum_u (\frac{\bar{\nu}_{uk}^{shp}}{\bar{\nu}_{uk}^{rte}} + \frac{\nu_{uk,t}^{shp}}{\nu_{uk,t}^{rte}})$
$\iota_{ik,t}$	Gamma	$\omega_{ik,t}^{shp}$ $\omega_{ik,t}^{rte}$	$a_\iota + a_\beta$ $a_\iota \frac{\mu_{ik,t}^{shp}}{\mu_{ik,t}^{rte}} + a_\beta \frac{\mu_{ik,t+1}^{shp}}{\mu_{ik,t+1}^{rte}}$

3 Batch Inference Algorithm

The mean-field Variational Inference of mGDMF as mentioned in Section 4 in the paper is shown below.

Algorithm 1 Batch Inference for mGDMF

```

1: Initialize  $\{\zeta, \rho, \kappa, \tau, \bar{\nu}, \bar{\mu}, \nu, \mu, \gamma, \omega, \phi\}$ .
2: Set  $K$ : # latent components.
3: Fix the parameters of the shape of Gamma distributions of  $\zeta, \rho, \gamma, \omega$ .
4: repeat
5:   for each time slice  $t = 1 \dots T$  do
6:     for each rating of user  $u$  to item  $i$  that  $y_{ui} \neq 0$  do
7:       Update the multinomial  $z_{ui,t}$ .
8:     end for
9:     for each user  $u$  do
10:      Update the static global factors  $\bar{\theta}_{uk}$ .
11:      Update rate of behavior  $\xi_u$ .
12:      for each user's attribute in metadata  $hu_m$  do
13:        Update rate of the weight.
14:      end for
15:      Update the shape and rate of dynamic correction factors  $\theta_{uk,t}$ .
16:      Update shape of user's auxiliary variables  $\lambda_{uk,t}$ .
17:    end for
18:    for each item  $i$  do
19:      Update the global factors  $\bar{\beta}_{ik}$ .
20:      Update rate of latent attractiveness  $\eta_i$ .
21:      for each item's attribute in metadata  $hi_n$  do
22:        Update rate of the weight.
23:      end for
24:      Update the shape and rate of dynamic correction factors  $\beta_{ik,t}$ .
25:      Update shape of item's auxiliary variables  $\lambda_{ik,t}$ .
26:    end for
27:  end for
28: until convergence

```

4 The Derivation of mGDMF

The mean-field family assumes each distribution is independent of each other and is governed by its own distribution.

$$\begin{aligned}
q(hu, hi, \xi, \eta, \bar{\theta}, \bar{\beta}, \lambda, \nu, \theta, \beta, z) = & \prod_m q(hu_m | \zeta_m) \prod_n q(hi_n | \rho_n) \prod_u q(\xi_u | \kappa_u) \prod_i q(\eta_i | \tau_i) \\
& \prod_{u,k} q(\bar{\theta}_{uk} | \bar{\nu}_{uk}) \prod_{i,k} q(\bar{\beta}_{ik} | \bar{\mu}_{ik}) \prod_{u,k,t} q(\theta_{uk,t} | \nu_{uk,t}) \prod_{i,k,t} q(\beta_{ik,t} | \mu_{ik,t}) \\
& \prod_{u,k,t} q(\lambda_{uk,t} | \gamma_{uk,t}) \prod_{i,k,t} q(\lambda_{ik,t} | \omega_{ik,t}) \prod_{u,i,t,k} q(z_{ui,t,k} | \phi_{ui,t,k})
\end{aligned} \tag{1}$$

4.1 Derivation of Metadata Integration

The derivation of the metadata integration is given here. With the Gamma distribution of user/item's metadata:

$$p(hu_m | a', b') \propto hu_m^{a'-1} \exp\{-b' * hu_m\} \tag{2}$$

$$p(hi_n | c', d') \propto hi_n^{c'-1} \exp\{-d' * hi_n\} \tag{3}$$

With the Gamma distribution of latent user preference and latent item attractiveness:

$$p(\xi_u|a, hu_m) \propto \left(\prod_{m=1}^M hu_m^{f_{u,m}a} \right) \exp\left\{-\left(\prod_{m=1}^M hu_m^{f_{u,m}} \right) \xi_u\right\} \quad (4)$$

$$p(\eta_i|c, hi_n) \propto \left(\prod_{n=1}^N hi_n^{f_{i,n}c} \right) \exp\left\{-\left(\prod_{n=1}^N hi_n^{f_{i,n}} \right) \eta_i\right\} \quad (5)$$

The posterior probability of weight hu_m and hi_n become:

$$p(hu_m|a', b', \xi_u) \propto p(hu_m|a', b') \prod_u p(\xi_u|a, hu_m) \propto hu_m^{a' + \aleph_m a - 1} \exp\left\{-(b' + \sum_u \xi_u) hu_m\right\} \quad (6)$$

$$p(hi_n|c', d', \eta_i) \propto p(hi_n|c', d') \prod_i p(\eta_i|c, hi_n) \propto hi_n^{c' + \chi_n c - 1} \exp\left\{-(d' + \sum_i \eta_i) hi_n\right\} \quad (7)$$

where \aleph_m is the number of users having attribute m and χ_n is the number of items that have attribute n . Thus, the posterior Gamma distribution of hu_m and hi_n are

$$hu_m|\xi_u \sim \text{Gamma}(a' + \aleph_m a, b' + \sum_u \xi_u) \quad (8)$$

$$hi_n|\eta_i \sim \text{Gamma}(c' + \chi_n c, d' + \sum_i \eta_i) \quad (9)$$

hu_m is affected by \aleph_m and the latent user preference (i.e., ξ_u) and hi_n is affected by χ_n and the latent item attractiveness (i.e., η_i)

We then update the variational *shape* (*shp*) and *rate* (*rte*) parameters for the user/item metadata.

$$(\zeta_m^{shp}, \zeta_m^{rte}) = (a' + \aleph_m a, b' + \sum_u \frac{\kappa_u^{shp}}{\kappa_u^{rte}}) \quad (10)$$

$$(\rho_n^{shp}, \rho_n^{rte}) = (c' + \chi_n c, d' + \sum_i \frac{\tau_i^{shp}}{\tau_i^{rte}}) \quad (11)$$

4.2 Derivation of User Latent Preference and Item Latent Attractiveness

With the Gamma distribution of latent user preference and latent item attractiveness:

$$p(\xi_u|a, hu_m) \propto \xi_u^{a-1} \exp\left\{-\left(\prod_{m=1}^M hu_m^{f_{u,m}} \right) \xi_u\right\} \quad (12)$$

$$p(\eta_i|c, hi_n) \propto \eta_i^{c-1} \exp\left\{-\left(\prod_{n=1}^N hi_n^{f_{i,n}} \right) \eta_i\right\} \quad (13)$$

With the Gamma distribution of user/item's global static factor:

$$p(\bar{\theta}_{uk}|b, \xi_u) \propto \xi_u^b \exp\{-\xi_u \bar{\theta}_{uk}\} \quad (14)$$

$$p(\bar{\beta}_{ik}|d, \eta_i) \propto \eta_i^d \exp\{-\eta_i \bar{\beta}_{ik}\} \quad (15)$$

The posterior probability of ξ_u and η_i becomes:

$$p(\xi_u|a, hu_m, \bar{\theta}_{uk}) \propto p(\xi_u|a, hu_m) \prod_k p(\bar{\theta}_{uk}|b, \xi_u) \propto \xi_u^{a+Kb-1} \exp\left\{-\left(\prod_{m=1}^M hu_m^{f_{u,m}} + \sum_k \bar{\theta}_{uk} \right) \xi_u\right\} \quad (16)$$

$$p(\eta_i | c, h_{i_n}, \bar{\beta}_{ik}) \propto p(\eta_i | c, h_{i_n}) \prod_k p(\bar{\beta}_{ik} | c, \eta_i) \propto \eta_i^{c+Kd-1} \exp\left\{-\left(\prod_{n=1}^N h_n^{f_{i,n}} + \sum_k \bar{\beta}_{ik}\right) \eta_i\right\} \quad (17)$$

where K is the number of latent components. Thus, the posterior Gamma distribution of ξ_u and η_i are

$$\xi_u | \bar{\theta}_{uk} \sim \text{Gamma}(a + Kb, \prod_{m=1}^M h_m^{f_{u,m}} + \sum_k \bar{\theta}_{uk}) \quad (18)$$

$$\eta_i | \bar{\beta}_{ik} \sim \text{Gamma}(c + Kd, \prod_{n=1}^N h_n^{f_{i,n}} + \sum_k \bar{\beta}_{ik}) \quad (19)$$

We then update the variational *shape* (*shp*) and *rate* (*rte*) parameters for the latent user preference and latent attractiveness:

$$(\kappa_u^{shp}, \kappa_u^{rte}) = (a + Kb, \prod_{m=1}^M (\frac{\zeta_m^{shp}}{\zeta_m^{rte}})^{f_{u,m}} + \sum_k \frac{\bar{\nu}_{uk}^{shp}}{\bar{\nu}_{uk}^{rte}}) \quad (20)$$

$$(\tau_i^{shp}, \tau_i^{rte}) = (c + Kd, \prod_{n=1}^N (\frac{\rho_n^{shp}}{\rho_n^{rte}})^{f_{i,n}} + \sum_k \frac{\bar{\mu}_{ik}^{shp}}{\bar{\mu}_{ik}^{rte}}) \quad (21)$$

4.3 Derivation of the Rating

To ensure the conjugacy of the model structure, inspired by [2, 1, 4, 3], the rating $y_{ui,t}$ is replaced with auxiliary latent variable $z_{ui,k,t} \sim \text{Poisson}((\theta_{uk,t} + \bar{\theta}_{uk})(\beta_{ik,t} + \bar{\beta}_{ik}))$. With the additive property of Poisson distribution, $y_{ui,t}$ is expressed as $y_{ui,t} = \sum_k z_{ui,k,t}$. The complete conditional for the auxiliary latent variable vector is

$$z_{ui,k,t} | \theta_{uk,t}, \bar{\theta}_{uk}, \beta_{ik,t}, \bar{\beta}_{ik} \sim \text{Mult}\left(y_{ui,t}, \frac{(\theta_{uk,t} + \bar{\theta}_{uk})(\beta_{ik,t} + \bar{\beta}_{ik})}{\sum_k (\theta_{uk,t} + \bar{\theta}_{uk})(\beta_{ik,t} + \bar{\beta}_{ik})}\right) \quad (22)$$

Thus, the update for the variational multinomial is

$$\begin{aligned} \phi_{ui,t,k} &= (\exp\{\Psi(\nu_{uk,t}^{shp}) - \log(\nu_{uk,t}^{rte})\} + \exp\{\Psi(\bar{\nu}_{uk}^{shp}) - \log(\bar{\nu}_{uk}^{rte})\}) \\ &\quad * (\exp\{\Psi(\mu_{ik,t}^{shp}) - \log(\mu_{ik,t}^{rte})\} + \exp\{\Psi(\bar{\mu}_{ik}^{shp}) - \log(\bar{\mu}_{ik}^{rte})\}) \end{aligned} \quad (23)$$

where $\Psi(\cdot)$ is the *digamma* function. This update comes from the expectation of the log of a Gamma variable, for example $E_q[\log(\theta_{uk,t})] = \Psi(\nu_{uk,t}^{shp}) - \log(\nu_{uk,t}^{rte})$.

4.4 Derivation of the Global Static Factor of User/Item

The global static factor of user/item is affected by the latent user preference/latent item attractiveness, the local dynamic factors and the observations with the conjugate Gamma-Poisson distribution. Thus, the complete conditional of those vectors are

$$\bar{\theta}_{uk} | \bar{\beta}_{ik}, \beta_{ik,t}, \xi_u, z_{ui,k,t} \sim \text{Gamma}(b + \sum_{i,t} z_{ui,k,t}, \xi_u + \sum_i (\bar{\beta}_{ik} + \sum_t \beta_{ik,t})) \quad (24)$$

$$\bar{\beta}_{ik} | \bar{\theta}_{uk}, \theta_{uk,t}, \eta_i, z_{ui,k,t} \sim \text{Gamma}(c + \sum_{u,t} z_{ui,k,t}, \eta_i + \sum_u (\bar{\theta}_{uk} + \sum_t \theta_{uk,t})) \quad (25)$$

The update for variational global static factor of user/item becomes

$$(\bar{\nu}_{uk}^{shp}, \bar{\nu}_{uk}^{rte}) = (b + \sum_{i,t} y_{ui,t} \phi_{ui,t,k}, \frac{\kappa_u^{shp}}{\kappa_u^{rte}} + \sum_i (\frac{\bar{\mu}_{ik}^{shp}}{\bar{\mu}_{ik}^{rte}} + \sum_t \frac{\mu_{ik,t}^{shp}}{\mu_{ik,t}^{rte}})) \quad (26)$$

$$(\bar{\mu}_{ik}^{shp}, \bar{\mu}_{ik}^{rte}) = (d + \sum_{u,t} y_{ui,t} \phi_{ui,t,k}, \frac{\tau_i^{shp}}{\tau_i^{rte}} + \sum_u (\frac{\bar{\nu}_{uk}^{shp}}{\bar{\nu}_{uk}^{rte}} + \sum_t \frac{\nu_{uk,t}^{shp}}{\nu_{uk,t}^{rte}})) \quad (27)$$

4.5 Derivation of the Local Dynamic Factor of User/Item

Initial states of the local dynamic factor of user/item are affected by the observations, the next auxiliary variables and the global static factors with the conjugate Gamma-Poisson distribution. Thus, the complete conditionals of those vectors are

$$\theta_{uk,1} | \bar{\beta}_{ik}, \beta_{ik,1}, \lambda_{uk,1}, z_{ui,k,1} \sim \text{Gamma}(a_\theta + a_\lambda + \sum_i z_{ui,k,1}, a_\theta b_\theta + a_\lambda \lambda_{uk,1} + \sum_i (\bar{\beta}_{ik} + \beta_{ik,1})) \quad (28)$$

$$\beta_{ik,1} | \bar{\theta}_{uk}, \theta_{uk,1}, \iota_{ik,1}, z_{ui,k,1} \sim \text{Gamma}(a_\beta + a_\iota + \sum_u z_{ui,k,1}, a_\beta b_\beta + a_\iota \iota_{ik,1} + \sum_u (\bar{\theta}_{uk} + \theta_{uk,1})) \quad (29)$$

The update of variational of initial states of local dynamic factor of user/item becomes:

$$(\nu_{uk,1}^{shp}, \nu_{uk,1}^{rte}) = (a_\theta + a_\lambda + \sum_i y_{ui,1} \phi_{ui,1,k}, a_\theta b_\theta + a_\lambda \frac{\gamma_{uk,1}^{shp}}{\gamma_{uk,1}^{rte}} + \sum_i (\frac{\bar{\mu}_{ik}^{shp}}{\bar{\mu}_{ik}^{rte}} + \frac{\mu_{ik,1}^{shp}}{\mu_{ik,1}^{rte}})) \quad (30)$$

$$(\mu_{ik,1}^{shp}, \mu_{ik,1}^{rte}) = (a_\beta + a_\iota + \sum_u y_{ui,1} \phi_{ui,1,k}, a_\beta b_\beta + a_\iota \frac{\omega_{ik,1}^{shp}}{\omega_{ik,1}^{rte}} + \sum_u (\frac{\bar{\nu}_{uk}^{shp}}{\bar{\nu}_{uk}^{rte}} + \frac{\nu_{uk,1}^{shp}}{\nu_{uk,1}^{rte}})) \quad (31)$$

Auxiliary variables of the local dynamic factor of user/item are affected by the previous local dynamic factor, the next local dynamic factor with the conjugate Gamma-Gamma distribution. Thus, the complete conditionals of those vectors are

$$\lambda_{uk,t} | \theta_{uk,t}, \theta_{uk,t+1} \sim \text{Gamma}(a_\lambda + a_\theta, a_\lambda \theta_{uk,t} + a_\theta \theta_{uk,t+1}) \quad (32)$$

$$\iota_{ik,t} | \beta_{ik,t}, \beta_{ik,t+1} \sim \text{Gamma}(a_\iota + a_\beta, a_\iota \beta_{ik,t} + a_\beta \beta_{ik,t+1}) \quad (33)$$

The update for variational of auxiliary variables of local dynamic factor of user/item becomes

$$(\gamma_{uk,t}^{shp}, \gamma_{uk,t}^{rte}) = (a_\lambda + a_\theta, a_\lambda \frac{\nu_{uk,t}^{shp}}{\nu_{uk,t}^{rte}} + a_\theta \frac{\nu_{uk,t+1}^{shp}}{\nu_{uk,t+1}^{rte}}) \quad (34)$$

$$(\omega_{ik,t}^{shp}, \omega_{ik,t}^{rte}) = (a_\iota + a_\beta, a_\iota \frac{\mu_{ik,t}^{shp}}{\mu_{ik,t}^{rte}} + a_\beta \frac{\mu_{ik,t+1}^{shp}}{\mu_{ik,t+1}^{rte}}) \quad (35)$$

Local dynamic factor of user/item at time $t > 1$ are affected by the observations, the previous auxiliary variables, the next auxiliary variables and the global static factors with the conjugate Gamma-Poisson distribution. Thus, the complete conditionals of those vector are

$$\theta_{uk,t} | \bar{\beta}_{ik}, \beta_{ik,t}, \lambda_{uk,1}, z_{ui,k,t} \sim \text{Gamma}(a_\theta + a_\lambda + \sum_i z_{ui,k,t}, a_\theta \lambda_{uk,t-1} + a_\lambda \lambda_{uk,t} + \sum_i (\bar{\beta}_{ik} + \beta_{ik,t})) \quad (36)$$

$$\beta_{ik,t} | \bar{\theta}_{uk}, \theta_{uk,t}, \iota_{ik,1}, z_{ui,k,t} \sim \text{Gamma}(a_\beta + a_\iota + \sum_u z_{ui,k,t}, a_\beta \iota_{ik,t-1} + a_\iota \iota_{ik,t} + \sum_u (\bar{\theta}_{uk} + \theta_{uk,t})) \quad (37)$$

The update for variational of local dynamic factor of user/item at time $t > 1$ becomes

$$(\nu_{uk,t}^{shp}, \nu_{uk,t}^{rte}) = (a_\theta + a_\lambda + \sum_i y_{ui,t} \phi_{ui,t,k}, a_\theta \frac{\gamma_{uk,t-1}^{shp}}{\gamma_{uk,t-1}^{rte}} + a_\lambda \frac{\gamma_{uk,t}^{shp}}{\gamma_{uk,t}^{rte}} + \sum_i (\frac{\bar{\mu}_{ik}^{shp}}{\bar{\mu}_{ik}^{rte}} + \frac{\mu_{ik,t}^{shp}}{\mu_{ik,t}^{rte}})) \quad (38)$$

$$(\mu_{ik,t}^{shp}, \mu_{ik,t}^{rte}) = (a_\beta + a_\iota + \sum_u y_{ui,t} \phi_{ui,t,k}, a_\beta \frac{\omega_{ik,t-1}^{shp}}{\omega_{ik,t-1}^{rte}} + a_\iota \frac{\omega_{ik,t}^{shp}}{\omega_{ik,t}^{rte}} + \sum_u (\frac{\bar{\nu}_{uk}^{shp}}{\bar{\nu}_{uk}^{rte}} + \frac{\nu_{uk,t}^{shp}}{\nu_{uk,t}^{rte}})) \quad (39)$$

References

- [1] D. B. Dunson and A. H. Herring. Bayesian latent variable models for mixed discrete outcomes. *Biostatistics*, 6(1):11–25, 2005.
- [2] P. Gopalan, J. M. Hofman, and D. M. Blei. Scalable recommendation with hierarchical Poisson factorization. In *UAI*, pages 326–335, 2015.
- [3] P. Gopalan, F. J. Ruiz, R. Ranganath, and D. Blei. Bayesian nonparametric Poisson factorization for recommendation systems. In *AISTATS*, pages 275–283, 2014.
- [4] M. Zhou, L. Hannah, D. B. Dunson, and L. Carin. Beta-negative binomial process and Poisson factor analysis. In *AISTATS*, pages 1462–1471, 2012.