Appendix for Thompson Sampling and Compositions in Knowledge Bases with Uncertainty

Abstract

This is the appendix of "Thompson Sampling and Compositions in Knowledge Bases with Uncertainty".

1. Posterior covariance of logistic output

Let R_k^* is a maximum a posterior solution of R_k given E. Then, the conditional posterior covariance of relation matrix R_k has the form of:

$$S_i^{-1} = \sum_{x_{ikj}} \sigma(e_i^\top R_k^* e_j)(1 - \sigma(e_i^{*\top} R_k^* e_j))\bar{e}_{ij}\bar{e}_{ij}^\top + I\sigma_r^{-1},$$

where $\bar{e}_{ij} = e_i \otimes e_j$.

2. Rao-Blackwellisation particle Thompson sampling

We also develop Rao-Blackwellisation particle Thompson sampling algorithm for the RESCAL with the Gaussian output model. The outline of the algorithm is described in Algorithm 1. With the Rao-Blackwellisation, we marginalise out the relation matrix R_k while computing the weight of each particle, but we still keep the same MCMC kernel to generate the samples. In theory, this will reduce the degeneracy problems for long running particles, but in our experiment, the difference between two models is not significant.

Algorithm 1 Rao-Blackwellised Particle Thompson Sampling for Gaussian output

Input: $\sigma_x, \sigma_e, \sigma_r$.
for $t = 1, 2, \ldots$ **do**
  Thompson Sampling:
  $h_t \sim \mathrm{Cat}(\mathbf{w}^{t-1})$
  $(i, k, j) \leftarrow \arg\max p(x_{ikj}|E^{h_t})$
  Query (i, k, j) and observe x_{ikj}
  $\mathcal{X}^t \leftarrow \mathcal{X}^{t-1} \cup x_{ikj}$
  Particle Filtering:
  $\forall h, w_h^t \propto p(x_{ikj}|E^{h_t})$ ▷ Reweighting
  if $\mathrm{ESS}(\mathbf{w}^t) \leq N$ **then**
    resample particles
    $w_h^t \leftarrow 1/H$
  end if
  for $h = 1$ **to** H **do**
    $\forall k, R_k^h \sim p(R_k|\mathcal{X}^t, E^{h_{t-1}})$ ▷ Auxiliary sampling
    $\forall i, e_i^h \sim p(e_i|\mathcal{X}^t, E_{-i}^h, \mathcal{R}^{h_t})$
  end for
end for

3. Compositional Relations

We provide the conditional posterior distribution of the compositional models.

3.1. Additive Compositionality

The conditional distribution of e_i given $E_{-i}, \mathcal{R}, \mathcal{X}^t, \mathcal{X}^{L(t)}$ is expanded from the posterior of BRESCAL by incorporating compositional triples.

$$p(e_i|E_{-i}, \mathcal{R}, \mathcal{X}^t, \mathcal{X}^{L(t)}) = \mathcal{N}(e_i|\mu_i, \Lambda_i^{-1}), \quad (1)$$

where

$$\begin{aligned}
\mu_i &= \Lambda_i^{-1}\xi_i \\
\Lambda_i &= \frac{1}{\sigma_x^2}\sum_{jk:x_{ikj}\in\mathcal{X}^t}(R_k e_j)(R_k e_j)^\top \\
&\quad + \frac{1}{\sigma_x^2}\sum_{jk:x_{jki}\in\mathcal{X}^t}(R_k^\top e_j)(R_k^\top e_j)^\top \\
&\quad + \frac{1}{\sigma_c^2}\sum_{jc:x_{icj}\in\mathcal{X}^{L(t)}}(R_c e_j)(R_c e_j)^\top \\
&\quad + \frac{1}{\sigma_c^2}\sum_{jc:x_{jci}\in\mathcal{X}^{L(t)}}(R_c^\top e_j)(R_c^\top e_j)^\top + \frac{1}{\sigma_e^2}I_D \\
\xi_i &= \frac{1}{\sigma_x^2}\sum_{jk:x_{ikj}\in\mathcal{X}^t}x_{ikj}R_k e_j + \frac{1}{\sigma_x^2}\sum_{jk:x_{jki}\in\mathcal{X}^t}x_{jki}R_k^\top e_j \\
&\quad + \frac{1}{\sigma_c^2}\sum_{jc:x_{icj}\in\mathcal{X}^{L(t)}}x_{icj}R_c e_j + \frac{1}{\sigma_c^2}\sum_{jc:x_{jci}\in\mathcal{X}^{L(t)}}x_{jci}R_c^\top e_j
\end{aligned}$$

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

The detail derivation of the posterior distribution is as follows:

$$
\begin{aligned}
&p(R_k|E,R_{-k},\mathcal{X}) \propto p(\mathcal{X}|R,E)p(R_k)\\
&\propto \prod_{x_{ikj}} \exp\Big\{-\frac{(x_{ikj}-e_i^\top R_k e_j)^2}{2\sigma_x^2}\Big\}\\
&\quad \prod_{x_{icj}} \exp\Big\{-\frac{(x_{icj}-e_i^\top R_c e_j)^2}{2\sigma_c^2}\Big\}\exp\Big\{-\frac{r_k^\top r_k}{2\sigma_r^2}\Big\}\\
&= \exp\Big\{-\frac{\sum_{x_{ikj}}(x_{ikj}-\bar{e}_{ij}^\top r_k)^2}{2\sigma_x^2}\\
&\quad -\frac{\sum_{x_{icj}}(x_{icj}-\bar{e}_{ij}^\top r_c)^2}{2\sigma_c^2}-\frac{r_k^\top r_k}{2\sigma_r^2}\Big\}\\
&= \exp\Big\{-\frac{\sum_{x_{ikj}}(x_{ikj}-\bar{e}_{ij}^\top r_k)^2}{2\sigma_x^2}\\
&\quad -\frac{\sum_{x_{icj}}(x_{icj}-\bar{e}_{ij}^\top(\frac{1}{|c|}r_k+\frac{|c|-1}{|c|})r_{c/k})^2}{2\sigma_c^2}-\frac{r_k^\top r_k}{2\sigma_r^2}\Big\}\\
&= \exp\Big\{-\frac{\sum_{x_{ikj}}-2x_{ikj}\bar{e}_{ij}^\top r_k + r_k^\top \bar{e}_{ij}\bar{e}_{ij}^\top r_k}{2\sigma_x^2}\\
&\quad -\frac{\sum_{x_{icj}}\frac{2}{|c|}r_k^\top(x_{icj}-\frac{(|c|-1)}{|c|}\bar{e}_{ij}^\top r_{c/k})+(\frac{1}{|c|^2}r_k^\top \bar{e}_{ij}\bar{e}_{ij}^\top r_k)}{2\sigma_c^2}\\
&\quad -\frac{r_k^\top r_k}{2\sigma_r^2}+const\Big\}\\
&\propto \exp\Big\{-\frac{1}{2}r_k^\top(\frac{1}{\sigma_x^2}\sum_{x_{ikj}}\bar{e}_{ij}\bar{e}_{ij}^\top+\frac{1}{|c|^2\sigma_c^2}\sum_{x_{icj}}\bar{e}_{ij}\bar{e}_{ij}^\top+\frac{1}{\sigma_r^2}I)r_k\\
&-r_k^\top\Big(\frac{\sum_{x_{ikj}}-x_{ikj}\bar{e}_{ij}}{\sigma_x^2}+\frac{\sum_{x_{icj}}|c|^{-1}(x_{icj}-\frac{(|c|-1)}{|c|}\bar{e}_{ij}^\top r_{c/k})}{\sigma_c^2}\Big)
\end{aligned}
$$

Completing the square results Equation 3.

To compute the conditional distribution of R_k, we first decompose R_c into two part where $R_c = \frac{1}{|c|}R_k + \frac{|c|-1}{|c|}R_{c/k}$, where $R_{c/k} = \sum_{k'\in c/k} R_{k'}$. The distribution of compositional triple is decomposed as follows:

$$x_{(i,c,l)} \sim \mathcal{N}(e_i^\top(\frac{1}{|c|}R_k + \frac{|c|-1}{|c|}R_{c/k})e_j, \sigma_c^2). \tag{2}$$

Then, the conditional distribution R_k given $R_{-k}, E, \mathcal{X}^t, \mathcal{X}^{L(t)}$ is

$$p(R_k|E,\mathcal{X}^t,\mathcal{X}^{L(t)},\sigma_r,\sigma_x) = \mathcal{N}(\text{vec}(R_k)|\mu_k,\Lambda_k^{-1}), \tag{3}$$

where

$$
\begin{aligned}
\mu_k &= \Lambda_k^{-1}\xi_k\\
\Lambda_k &= \frac{1}{\sigma_x^2}\sum_{ij:x_{ikj}\in\mathcal{X}^t}\bar{e}_{ij}\bar{e}_{ij}^\top + \frac{1}{\sigma_r^2}I_{D^2}\\
&\quad + \frac{1}{|c|^2\sigma_c^2}\sum_{ij:x_{icj}\in\mathcal{X}^{L(t)},\,k\in c}\bar{e}_{ij}\bar{e}_{ij}^\top\\
\xi_k &= \frac{1}{\sigma_x^2}\sum_{ij:x_{ikj}\in\mathcal{X}^t}x_{ikj}\bar{e}_{ij}\\
&\quad + \frac{1}{|c|\sigma_c^2}\sum_{ij:x_{icj}\in\mathcal{X}^{L(t)},\,k\in c}x_{icj}\bar{e}_{ij} - \frac{|c|-1}{|c|}\bar{e}_{ij}r_{c/k}^\top\bar{e}_{ij}\\
\bar{e}_{ij} &= e_i\otimes e_j.
\end{aligned}
$$

Vectorisation of R_c and $R_{c/k}$ are represented as r_c and $r_{c/k}$, respectively.

3.2. Multiplicative Compositionality

Given a sequence of relations including relation k, R_k is placed in the middle of the compositional sequence, i.e., $e_i^\top R_{c(1)}R_{c(2)}\dots R_{c(\delta_k)}\dots R_{c(|c|-1)}R_{c(|c|)}e_j$, where δ_k is the index of relation k. For notational simplicity, we will denote the left side $e_i^\top R_{c(1)}R_{c(2)}\dots R_{c(\delta_k-1)}$ as $\bar{e}_{ic(:\delta_k)}^\top$, and the right side $R_{c(\delta_k+1)}\dots R_{c(|c|-1)}R_{c(|c|)}e_j$ as $\bar{e}_{ic(\delta_k:)}$, therefore we can rewrite the mean parameter as $\bar{e}_{ic(:\delta_k)}^\top R_k \bar{e}_{ic(\delta_k:)}$. With the simplified notations, the conditional of R_k is

$$p(R_k|E,\mathcal{X},\sigma_r,\sigma_x) = \mathcal{N}(\text{vec}(R_k)|\mu_k,\Lambda_k^{-1}), \tag{4}$$

where

$$
\begin{aligned}
\mu_k &= \Lambda_k^{-1}\xi_k\\
\Lambda_k &= \frac{1}{\sigma_x^2}\sum_{ij:x_{ikj}\in\mathcal{X}^t}(e_i\otimes e_j)(e_i\otimes e_j)^\top + \frac{1}{\sigma_r^2}I_{D^2}\\
&+\frac{1}{\sigma_c^2}\sum_{ij:x_{icj}\in\mathcal{X}^{L(t)},\,k\in c}(\bar{e}_{ic(:\delta_k)}\otimes\bar{e}_{jc(\delta_k:)})(\bar{e}_{ic(:\delta_k)}\otimes\bar{e}_{jc(\delta_k:)})^\top\\
\xi_k &= \frac{1}{\sigma_x^2}\sum_{ij:x_{ikj}\in\mathcal{X}^t}x_{ikj}(e_j\otimes e_i)\\
&\quad + \frac{1}{\sigma_c^2}\sum_{ij:x_{icj}\in\mathcal{X}^{L(t)},\,k\in c}x_{icj}(\bar{e}_{ic(:\delta_k)}\otimes\bar{e}_{jc(\delta_k:)}).
\end{aligned}
$$

The conditional distribution of e_i given the rest is the same as the additive compositional case.