Technical Report: Variational forms and updates for the WARBLE model

Joan Capdevila[1,2], Jesús Cerquides[3], and Jordi Torres[1,2]

[1]Department of Computer Architecture, UPC BarcelonaTech
[2]Department of Computer Science, Barcelona Supercomputing Center (BSC-CNS)
[3]AI Research Institute (IIIA), Spanish National Research Council (CSIC)

April 2016

1 Introduction

This document summarizes the variational inference updates within the mean-field approximation of the WARBLE model depicted in Fig. 1.

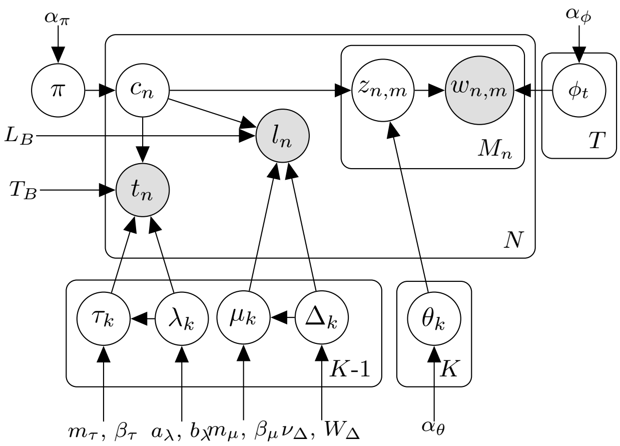

Figure 1: The Probabilistic Graphical Model for WARBLE

The mean-field approximation considers the following factorized distribution $q(X;\eta)$,

$$q(X;\eta) = q(\pi)\prod_{t=1}^{T} q(\phi_t)\prod_{n=1}^{N} q(c_n)\prod_{m=1}^{M_n} q(z_{n,m})$$
$$q(\theta_K)\prod_{k=1}^{K-1} q(\tau_k)q(\lambda_k)q(\mu_k)q(\Delta_k)q(\theta_k) \tag{1}$$

where X stands for the set of random variables containing c, z, π, τ, λ, μ, Δ, θ and ϕ and η the variational parameters to be detailed later.

The goal of variational inference is to minimize the Kullback-Leibler (KL) divergence between the posterior distribution $p(X|D;\Gamma)$ (where D stands for our data, namely l, t, and w) and the factorized distribution $q(X;\eta)$.

$$\underset{q(X;\eta)}{\operatorname{argmin}}\ \mathbb{KL}(q(X;\eta), p(X|D;\Gamma)) \tag{2}$$

where Γ stands for the model hyperparameters L_B, T_B, α_π, α_θ, α_ϕ, m_τ, β_τ, a_λ, b_λ, m_μ, β_μ, ν_Δ and W_Δ.

However, the mean-field variational update for a random variable x whatsoever can be expressed as

$$q(x) \propto exp\left(\int_{X-x} q(X;\eta)\log p(X,D;\Gamma)\right) \tag{3}$$

where $\log p(X, D; \Gamma)$ is the logarithm of the join probability distribution for the WARBLE model in Fig. 1

2 Functional forms and variational parameters

In this section we present the functional forms and the variational parameters for each variable X, which have been derived through Eq. 3

2.1 The mixture proportions distribution $q(\pi)$

The mixture proportions approximate distribution results in a Dirichlet with parameters π'_k.

$$q(\pi) \sim \; Dir(\pi|\pi'_k)$$
$$\pi'_k = \alpha_\pi + \sum_{n=1}^{N} c'_{n,k} = \alpha_\pi + N_{c_k}$$

2.2 The mixture assignments distribution $q(c_n)$

Consider the following expressions:

$$\mathbb{E}(\log \pi)_k = \int_\pi q(\pi; \pi'_k) \log \pi = \Psi(\pi'_k) - \Psi(\sum_{k=1}^{K} \pi'_k)$$
$$\mathbb{E}(\log \theta_k)_t = \int_{\theta_k} q(\theta_k; \theta'_{k,t}) \log \theta_k = \Psi(\theta'_{k,t}) - \Psi(\sum_{t=1}^{T} \theta'_{k,t})$$

where $\Psi(.)$ corresponds to the digamma function.

The mixture assignments distribution is a Categorical with parameters c'_{nk}.

$$q(c_n) \sim \; Cat(c_n|c'_{nk})$$
$$\tilde{c}'_{nk} \propto exp\Bigg(\mathbb{E}(\log \pi)_k + \sum_{m=1}^{M_n} w_{n,m} \sum_{t=1}^{T} z_{n,m,t}\, \mathbb{E}(\log \theta_k)_t$$
$$+\, \mathbb{I}(k = K)\, (\log Hist(l_n; L_B) + \log Hist(t_n; T_B))$$
$$+\, \mathbb{I}(k \neq K) \Bigg(-\log 2\pi + \frac{1}{2}\Bigg(2 \log 2 + \Psi(\frac{\nu'_k}{2}) + \Psi(\frac{\nu'_k - 1}{2})$$
$$+\; \log |W'_k| - \frac{2}{\beta'_{\mu_k}} - \nu'_k(l_n - m'_{\mu_k})^T W_k(l_n - m'_{\mu_k})\Bigg)$$
$$-\frac{1}{2}\log 2\pi + \frac{1}{2}\left(\Psi(a'_k) - \log b'_k - \frac{1}{\beta'_{\tau_k}} - \frac{a'_k}{b'_k}(t_n - m'_{\tau_k})^2\right)\Bigg)$$
$$c'_{nk} = \frac{\tilde{c}'_{nk}}{\sum_{k=1}^{K} \tilde{c}'_{nk}}$$

2.3 The topic assignments distributions $q(z_{n,m})$

Consider the following expression:

$$\mathbb{E}(\log \phi_t)_v = \int_{\phi_t} q(\phi_t; \phi'_{t,v}) \log \phi_t = \Psi(\phi'_{t,v}) - \Psi(\sum_{v=1}^{V} \phi'_{t,v})$$

The topic assignments distribution is a Categorical with parameters $z'_{n,m,t}$.

$$q(z_{n,m}) \sim \; Cat(z_{n,m}|z'_{n,m,t})$$
$$\tilde{z}'_{n,m,t} \propto exp\left(\mathbb{E}(\log \phi_t)_m + \sum_{k=1}^{K} c'_{nk}\, \mathbb{E}(\log \theta_k)_t\right)$$
$$z'_{n,m,t} = \frac{\tilde{z}'_{n,m,t}}{\sum_{t=1}^{T} \tilde{z}'_{n,m,t}}$$

2.4 The word distributions $q(\phi_t)$

The word distribution for each topic t is Dirichlet with parameters ϕ'_t.

$$q(\phi_t) \sim \; Dir(\phi_t|\phi'_t)$$
$$\phi'_t = \alpha_\phi + \sum_{n=1}^{N} w_{n,m} z'_{n,m,t}$$

2.5 The temporal mean and variance distributions $q(\tau_k)$, $q(\lambda_k)$

The temporal mean and variance distributions for each component k are Normal and Gamma with parameters m'_{τ_k}, β'_{τ_k} and a'_{λ_k}, b'_{λ_k} respectively.

$$q(\tau_k) \sim \; N(\tau_k|m'_{\tau_k}, \beta'_{\tau_k}\frac{a'_\lambda}{b'_\lambda})$$
$$m'_{\tau_k} = \frac{m_\tau \beta_\tau + \sum_{n=1}^{N} c'_{nk} t_n}{\beta'_{\tau_k}}$$
$$\beta'_{\tau_k} = \beta_\tau + N_{c_k}$$
$$q(\lambda_k) \sim \; G(\lambda_k|a'_{\lambda_k}, b'_{\lambda_k})$$
$$a'_{\lambda_k} = a_\lambda + \frac{N_{c_k}}{2}$$
$$b'_{\lambda_k} = b_\lambda + \frac{1}{2}\sum_{n=1}^{N}(t_n - m'_{\tau_k})^2 + \frac{\beta_\tau}{2}(m'_{\tau_k} - m_\tau)^2$$

2.6 The spatial mean and variance distributions $q(\mu_k)$, $q(\Delta_k)$

Consider the following expressions:

$$\overline{l_k} = \frac{1}{N_{c_k}} \sum_{n=1}^{N} c'_{n,k} l_n$$

$$S_k = \frac{1}{N_{c_k}} \sum_{n=1}^{N} c'_{n,k} (l_n - \overline{l_k})^T (l_n - \overline{l_k})$$

The spatial mean and variance distributions for each component k are Normal and Wishart with parameters m'_{μ_k}, β'_{μ_k} and ν'_k, W'_k, respectively.

$$\begin{aligned}
q(\mu_k) &\sim \ N(\mu_k | m_{\mu'_k}, \beta'_{\mu_k} \nu'_k W'_k) \\
m'_{\mu_k} &= \frac{m_\mu \beta_\mu + N_{c_k} \overline{l_k}}{\beta'_{\mu_k}} \\
\beta'_{\mu_k} &= \beta_\mu + N_{c_k} \\
q(\Delta_k) &\sim \ W(\Delta_k | \nu'_k, W'_k) \\
\nu'_k &= \nu_\Delta + N_{c_k} \\
W'_k &= \left(W_\Delta^{-1} + N_{c_k} S_k + \frac{\beta_\mu N_{c_k}}{\beta_\mu + N_{c_k}} (\overline{l_k} - m_\mu)^T (\overline{l_k} - m_\mu) \right)^{-1}
\end{aligned}$$

2.7 The topic distributions $q(\theta_k)$

The topic distributions for each component k are Dirichlet with parameters θ'_k.

$$\begin{aligned}
q(\theta_k) &\sim \ Dir(\theta_k | \theta'_k) \\
\theta'_k &= \alpha_\theta + \sum_{n=1}^{N} c'_{nk} \sum_{m=1}^{M_n} w_{n,m} z'_{n,m,t}
\end{aligned}$$