Supplementary Section: Social Role Discovery in Human Events

Vignesh Ramanathan*, Bangpeng Yao†, Li Fei-Fei†
*Department of Electrical Engineering, †Department of Computer Science
Stanford University
{vigneshr, bangpeng.yao, feifeili}@cs.stanford.edu

A. Update equations for variational inference

A lower bound on the log likelihood of the CRF can be derived using Jensen's inequality as shown in Eq. 1.

$$\begin{aligned} \log p\left(s_E, \alpha, \beta | \Sigma_\alpha, \Sigma_\beta\right) &\geq \mathbb{L}(q, \Psi_u, \Psi_p) \\ &= E_q\left[\log p(\alpha, \beta|\Sigma_\alpha, \Sigma_\beta)\right] + \\ &\quad \sum_v E_q\left[\log p(s^v|\alpha, \beta)\right] + H, \end{aligned} \tag{1}$$

where s^v is the complete role assignment to all people in the video v, and H is the entropy of the variational distribution q shown in Eq. 2.

$$\begin{aligned} & q(\alpha, \beta, s_E|\lambda_\alpha, \lambda_\beta, \sigma_\alpha^2, \sigma_\beta^2, \phi, \psi) = \\ & \prod_j q(\alpha^j|\lambda_{\alpha^j}, \sigma^2_{\alpha^j}) \prod_k q(\beta^k|\lambda_{\beta^k}, \sigma^2_{\beta^k}) \prod_v q(s^v|\phi^v, \psi^v) \end{aligned} \tag{2}$$

Now, $E_q\left[\log p(s^v|\alpha, \beta)\right]$ in Eq. 1 can be expanded as

$$\begin{aligned} E_q\left[\log p(s^v|\alpha, \beta)\right] = \lambda_\alpha \cdot E_q\left[\Psi_u(s^v)\right] + \\ \lambda_\beta \cdot E_q\left[\Psi_p(s^v)\right] - E_q\left[Z_v\right], \end{aligned} \tag{3}$$

where, $Z_v = \log\left\{\sum_{s^v} \exp\left(\alpha \cdot \Psi_u(s^v) + \beta \cdot \Psi_p(s^v)\right)\right\}$ is the log partition function. Using the fact, $\log x \leq a^{-1}x - 1 + \log a$, we can establish a lower bound on $E_q\left[\log p(s^v|\alpha, \beta)\right]$ as shown below

$$\begin{aligned} E_q\left[\log p(s^v|\alpha, \beta)\right] \geq \lambda_\alpha \cdot E_q\left[\Psi_u(s^v)\right] + \\ \lambda_\beta \cdot E_q\left[\Psi_p(s^v)\right] - \frac{h_v(q)}{\zeta_v} - log(\zeta_v), \end{aligned} \tag{4}$$

where ζ_v is a variational parameter and $h_v(q)$ is defined as

$$\begin{aligned} h_v(q) &= \sum_{s^v} E_q\left[\exp\left\{\alpha \cdot \Psi_u(s^v) + \beta \cdot \Psi_p(s^v)\right\}\right] \\ &= \sum_{s^v} \exp\left\{\sum_{p_i^v} \lambda_\alpha \cdot \Psi_u(p_i^v, s_i^v) + \frac{\sigma_\alpha^2}{2} \cdot \Psi_u^2(p_i^v, s_i^v)\right. \\ &\quad \left. + \sum_{p_j^v \neq p_m^v} \lambda_\beta \cdot \Psi_p(p_m^v, p_j^v, s_j^v) + \frac{\sigma_\beta^2}{2} \cdot \Psi_p^2(p_m^v, p_j^v, s_j^v)\right\} \end{aligned} \tag{5}$$

Given $\Sigma_\alpha, \Sigma_\beta$, we update the parameters through a coordinate ascent method to maximize the lower bound in Eq. 1. ζ_v is updated to $h_v(q)$ at each iteration. The closed form update equations for $\phi^v(p_i^v), \psi^v_{(i)}(p_j^v, s)$ at each iteration are shown in Eq. 6.

$$\begin{aligned} \phi^v(p_i^v) &\propto \exp\left\{\lambda_{\alpha_{m_E}} \cdot \Psi_u(p_i^v) + \right. \\ &\quad \left. \sum_{j \neq i} \sum_{s \neq m_E} \psi_{(i)}(p_j^v, s)\left[\lambda_{\beta_s} \cdot \Psi_p(p_j^v, p_i^v)\right]\right\} \\ \psi^v_{(i)}(p_j^v, s) &\propto \exp\left\{\phi^v(p_i^v)\lambda_{\alpha_s} \cdot \Psi_u(p_j^v)\right. \\ &\quad \left. + \phi^v(p_i^v)\left[\lambda_{\beta_s} \cdot \Psi_p(p_i^v, p_j^v)\right]\right\} \end{aligned} \tag{6}$$

At each iteration, the values of $\lambda_\alpha, \lambda_\beta$, and $\sigma_\alpha^2, \sigma_\beta^2$ are updated using L-BFGS. The gradients of $\mathbb{L}$ with respect to λ_{α_s} and λ_{β_s} are given below

$$\begin{aligned} \nabla_{\lambda_{\alpha_s}} \mathbb{L} &= \Sigma_{\alpha_s}^{-1}\lambda_{\alpha_s} + \sum_v \left\{\sum_i E_q[\Psi_u(p_i^v, s)] - \right. \\ &\quad \left. \zeta_v^{-1} \nabla_{\lambda_{\alpha_s}} h_v(q)\right\} \\ \nabla_{\lambda_{\beta_s}} \mathbb{L} &= \Sigma_{\beta_s}^{-1}\lambda_{\beta_s} + \sum_v \left\{\sum_{\substack{i,j \\ j \neq i}} E_q[\Psi_p(p_i^v, p_j^v, s)] - \right. \\ &\quad \left. \zeta_v^{-1} \nabla_{\lambda_{\alpha_s}} h_v(q)\right\}, \end{aligned} \tag{7}$$

where $\Sigma_{\alpha_s}, \Sigma_{\beta_s}$ are the components of $\Sigma_\alpha, \Sigma_\beta$ corresponding to α_s, β_s respectively. As before, $\Psi_p(p_i^v, p_j^v, s)$ is the pairwise feature when p_i^v is the reference role and p_j^v is assigned the role s. The gradients of $\mathbb{L}$ with respect to $\sigma^2_{\alpha^k}$ and $\sigma^2_{\beta^k}$ are given below

$$\nabla_{\sigma^2_{\alpha^k}} \mathbb{L} = -\frac{1}{2}\Sigma^{-1}_{\alpha^k} - \sum_v \zeta_v^{-1} \nabla_{\sigma^2_{\alpha^k}} h_v(q) + \frac{1}{2\sigma^2_{\alpha^k}} \tag{8}$$

$$\nabla_{\sigma^2_{\beta^k}} \mathbb{L} = -\frac{1}{2}\Sigma^{-1}_{\beta^k} - \sum_v \zeta_v^{-1} \nabla_{\sigma^2_{\beta^k}} h_v(q) + \frac{1}{2\sigma^2_{\beta^k}},$$

where Σ_{α^k} and Σ_{β^k} are the k^{th} diagonal elements in Σ_α and Σ_β respectively.

It is to be noted that the assumption of significant interaction only with the reference role, helps us exactly compute $h_v(q), \nabla_\lambda h_v(q), \nabla_{\sigma^2} h_v(q)$ through a clique-tree messgae passing algorithm. The exact computation of $h_v(q)$ is intractable in a fully connected graph with interaction among all social roles.

Implementation details ζ_v is initialized to $1E6$ in our experiments. Also, the hyperparameters Σ_α and Σ_β are diagonal matrices whose non-zero entries are all set to $0.01, 0.1$, or 10 based on the variance of the unary and pairwise features. This is indicative of the amount of variance in the respective features.

B. Optimization for role assignment from probabilities

In every video v, the person p_m^v with the highest value of $\phi^v(p_m^v)$ is assigned the reference role. The corresponding variational probability $\psi^v_{(m)}$ is then used to assign secondary roles to other people in the video. While assigning secondary roles, we enforce a lower l and upper u bound on the number of people assigned a secondary role s in the event. Let $\mathbb{P}_{-m_E}$ be the set of people not assigned the reference role in event E. Let ψ be the secondary role probability matrix, where each row corresponds to a person $p_k \in \mathbb{P}_{-m_E}$ and each column represents a secondary role s. ψ can be obtained by stacking $\psi^v_{(m)}$ from all videos. Y is the secondary role assignment matrix with same dimensions as ψ, where an entry Y_{ks} is set to 1 if the person p_k is assigned the role s. Secondary role assignment is then carried out by solving the linear integer program in Eq. 9 to maximize the probability of role assignment under given constraints.

$$\begin{aligned} \max_{Y} \quad & Trace\left(Y^T\psi\right), \\ \text{subject to} \quad & Y\mathbf{1} = \mathbf{1}, \\ & l \le Y^T\mathbf{1} \le u, \\ & Y_{ks} \in \{0,1\} \, \forall \, p_k \in \mathbb{P}_{-m_E} \end{aligned} \tag{9}$$

We enforce the same constraints in our baseline models as well.