Supplementary note for "Transformation of Markov Random Fields for Marginal Distribution Estimation"

Masaki Saito Takayuki Okatani
Tohoku University, Japan
{msaito, okatani}@vision.is.tohoku.ac.jp

In this note, we show the details of mathematical derivations omitted in our main paper.

A. Derivation of the MRF transformation (Proof of Theorem 3.2)

Section 3.3 presents the proposed method for MRF transformation. As explained there, the derivation follows the standard methodology. Namely, as it is difficult to compute the marginal distributions from $p_0(\boldsymbol{x})$, choosing its approximate distribution $q_0(\boldsymbol{x})$ such that the marginal distributions are easier to compute, we make it as close to p_0 as possible. This is done by minimizing the KL distance $\mathcal{D}[q_0 \| p_0]$, which reduces to that of the following free energy:

$$\mathcal{F}[q_0] = \langle E_0(\boldsymbol{x}) \rangle_{q_0} - \mathcal{H}[q_0], \tag{31}$$

where

$$\langle E_0(\boldsymbol{x}) \rangle_{q_0} = \sum_{\boldsymbol{x}} q_0(\boldsymbol{x}) E_0(\boldsymbol{x}),$$

and

$$\mathcal{H}[q_0] = -\sum_{\boldsymbol{x}} q_0(\boldsymbol{x}) \ln q_0(\boldsymbol{x}).$$

The main idea of our method is to choose $q_0(\boldsymbol{x})$ as

$$q_0(\boldsymbol{x}) = \sum_{\boldsymbol{z}_1} q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1) q_1(\boldsymbol{z}_1). \tag{32}$$

As is explained in the main paper, we choose $q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1)$ arbitrarily depending on applications. Then, the problem turns to the estimation of $q_1(\boldsymbol{z}_1)$.

To do so, we rewrite Eq.(31) by using Eq.(32). The first term of Eq.(31) is rewritten as

$$\begin{aligned}\langle E_0(\boldsymbol{x}) \rangle_{q_0} &= \sum_{\boldsymbol{z}_1} q_1(\boldsymbol{z}_1) \sum_{\boldsymbol{x}} q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1) E_0(\boldsymbol{x}) \\ &= \left\langle \sum_{\boldsymbol{x}} q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1) E_0(\boldsymbol{x}) \right\rangle_{q_1}. \end{aligned} \tag{33}$$

To rewrite the second term of Eq.(31), we use $q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}) = q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1) q_1(\boldsymbol{z}_1)/q_0(\boldsymbol{x})$, which can be rewritten as

$$\ln q_0(\boldsymbol{x}) = \ln q(\boldsymbol{z}_1) + \ln q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1) - \ln q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}). \tag{34}$$

Note that this equation holds true for any $\boldsymbol{z}_1 \in \mathcal{Z}_1$, where $\mathcal{Z}_1$ is an appropriately defined variable space of $\boldsymbol{z}_1$. Using Eq.(34), we can rewrite $\mathcal{H}[q_0]$ in Eq.(31) as

$$\begin{aligned}\mathcal{H}[q_0] = &-\sum_{\boldsymbol{z}_1} q_1(\boldsymbol{z}_1) \ln q_1(\boldsymbol{z}_1) \\ &- \sum_{\boldsymbol{z}_1} q_1(\boldsymbol{z}_1) \sum_{\boldsymbol{x}} q(\boldsymbol{x}|\boldsymbol{z}_1) \ln q(\boldsymbol{x}|\boldsymbol{z}_1) \\ &+ \sum_{\boldsymbol{x}} q_0(\boldsymbol{x}_0) \sum_{\boldsymbol{z}_1} q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}) \ln q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}). \end{aligned} \tag{35}$$

The substitution of Eq.(33) and Eq.(35) into Eq.(31) yields

$$\begin{aligned}\mathcal{F}[q_0] = &\left\langle \sum_{\boldsymbol{x}} q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1)(E_0(\boldsymbol{x}) + \ln q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1)) \right\rangle_{q_1} \\ &+ \sum_{\boldsymbol{z}_1} q_1(\boldsymbol{z}_1) \ln q_1(\boldsymbol{z}_1) \\ &+ \left\langle \sum_{\boldsymbol{z}_1} q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}) \ln q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}) \right\rangle_{q_0}. \end{aligned} \tag{36}$$

The second term of the right hand side is the entropy of q_1, which we write as $\mathcal{H}[q_1]$. Defining $E_1(\boldsymbol{z}_1)$ and $S_1(\boldsymbol{x})$ as

$$E_1(\boldsymbol{z}_1) = \sum_{\boldsymbol{x}} q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1) \left\{ E_0(\boldsymbol{x}) + \ln q_{0,1}(\boldsymbol{x}|\boldsymbol{z}_1) \right\} \tag{37}$$

and

$$S_1(\boldsymbol{x}) = -\sum_{\boldsymbol{z}} q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}) \ln q_{0,1}(\boldsymbol{z}_1|\boldsymbol{x}), \tag{38}$$

respectively, we can rewrite $\mathcal{F}[q_0]$ as

$$\mathcal{F}[q_0] = \langle E_1(\boldsymbol{z}_1) \rangle_{q_1} - \mathcal{H}[q_1] + \langle S_1(\boldsymbol{x}) \rangle_{q_0}. \tag{39}$$

Lemma 3.1 states that the third term vanishes when the condition given in Lemma 3.1 is met. This is self-evident from Eq.(38). We have thus proved Theorem 3.2.

B. Downsizing a CRF for semantic labeling

In Section 6.2, we show the experiments of downsizing a CRF for semantic labeling. We considered a grid CRF whose energy is given by

$$E(\boldsymbol{x}|\mathcal{I};\theta) = \sum_i f_i(x_i|\mathcal{I}) \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} \theta_{st}\delta(x_i - s)\delta(x_j - t), \quad (40)$$

where θ_{st} is the parameter representing the interaction between the label s and t, which we want to determine through learning. To be specific, we determine $\theta = \{\theta_{st}\}$ by minimizing the negative log-likelihood

$$J(\theta) = -\frac{1}{M}\sum_m \ln p(\boldsymbol{x}^m|\mathcal{I}^m;\theta). \quad (41)$$

As mentioned in Section 6.2, we use the (stochastic) gradient descent method for the minimization. The gradient of $J(\theta)$ is given as

$$\frac{\partial J}{\partial\theta_{st}} = \frac{1}{M}\sum_m \sum_{(i,j)\in\mathcal{E}} \delta(x_i^m - s)\delta(x_j^m - t) - \frac{1}{M}\sum_m \sum_{(i,j)\in\mathcal{E}} p_{ij}(s,t|\mathcal{I}^m;\theta), \quad (42)$$

where $p_{ij}(x_i, x_j|\mathcal{I}^m;\theta)$ is the marginal distribution between the i-th and j-th sites. In our experiments, we choose BP for their estimation. In this configuration, we examine the effectiveness of the proposed MRF transformation.

B.1. Grouping of discrete labels

Let $q_{ij}^1(z_i, z_j|\mathcal{I}^m;\theta)$ be the marginal distribution of the transformed CRF, which is estimated by using the augmented energy of Eq.(20), and let $q_{ij}^0(x_i, x_j|\mathcal{I}^m;\theta)$ be the marginal distribution of the original energy function. Using Eqs.(14), (19), and (20), these two are related as

$$q_{ij}^0(x_i, x_j|\mathcal{I}^m;\theta) = \frac{1}{|\mathcal{X}_i^u||\mathcal{X}_j^v|} q_{ij}^1(u, v|\mathcal{I}^m;\theta), \quad (43)$$

where u and v on the right hand side are the labels (indeces) of the supports $\mathcal{X}_i^u$ and $\mathcal{X}_j^v$ within which x_i and x_j lie, respectively; that is, $x_i \in \mathcal{X}_i^u$ and $x_j \in \mathcal{X}_j^v$.

B.2. Coarse graining of MRFs

Similarly, using Eqs.(21), (24), and (25), we can express q_{ij}^0 with q_{ij}^1. If $(i,j) \in \mathrm{In}(k)$, q_{ij}^0 can be expressed as

$$q_{ij}^0(x_i, x_j|\mathcal{I}^m;\theta) = \delta(x_i - x_j) q_k^1(x_i|\mathcal{I}^m;\theta), \quad (44)$$

where $q_k^1(x_i|\mathcal{I}^m;\theta)$ is the marginal distribution of the i-th site estimated from $q^1(\boldsymbol{z}_1)$. If $(i,j) \in \mathrm{Ex}(k,l)$, q_{ij}^0 is expressed as

$$q_{ij}^0(x_i, x_j|\mathcal{I}^m;\theta) = q_{k,l}^1(x_i, x_j|\mathcal{I}^m;\theta), \quad (45)$$

where $q_{k,l}^1(x_i, x_j)$ is the marginal distribution estimated from $q^1(\boldsymbol{z}_1)$. Thus, we can regard $q_{k,l}^1$ as q_{ij}^0 in this case.