Title of the presentation]CSE-5830: Project Proposal

1 Efficient Inference of fully connected CRF

1.1 Definition of Conditional Random Fields

A conditional random field $(\mathbf{I}, \mathbf{X})$ is characterized by a Gibbs distribution

$$P(\mathbf{X}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})} \exp\left(- \sum_{c \in C_{\mathcal{G}}} \phi_c(\mathbf{X}_c|\mathbf{I}) \right),$$

where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a graph on $\mathbf{X}$ and each clique c in a set of cliques $C_{\mathcal{G}}$ in $\mathcal{G}$ induces a potential ϕ_c. The Gibbs energy of a labeling $\mathbf{x} \in \mathcal{L}^N$ is

$$E(\mathbf{x}|\mathbf{I}) = \sum_{c \in C_{\mathcal{G}}} \phi_c(\mathbf{x}_c|\mathbf{I}).$$

In the fully connected pairwise CRF model, $\mathcal{G}$ is the complete graph on $\mathbf{X}$ and $C_{\mathcal{G}}$ is the set of all unary and pairwise cliques. The corresponding Gibbs energy i :

$$E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j),$$

1.2 Mean Field Approximation of Inference of CRF

Instead of computing the exact distribution $P(\mathbf{X})$, the mean field approximation would transfer the computations of CRF into a distribution $Q(\mathbf{X})$ with simpler structure. The ideal choice of the approximation measure should satisfy

- Q can be expressed product of independent marginals, $Q(\mathbf{X}) = \prod_i Q_i(X_i)$.
- Q will minimize the KL-divergence $\mathcal{D}(Q\|P)$.

By applying Lagrainge multiplier method, we could obtain the necessary condition of the optimal measure is a system of nonlinear equations:

$$Q_i(x_i = l) = \frac{1}{Z_i} \exp\left\{ -\psi_u(x_i) - \sum_{l' \in \mathcal{L}} \mu(l, l') \sum_{m=1}^{K} w^{(m)} \sum_{j \neq i} k^{(m)}(f_i, f_j) Q_j(l') \right\}.$$

Recall that we have N different data samples and for each sample we have $L-1$ parameters to be determinated, *i.e.*$\{Q_i(x_i = l)\}_{l=1}^{L-1}$. Thus analytically solving this system of equations may still be hard. Practically, a fixed point iteration approximation could be an efficient way to obtain the numerical solution. The fixed -point iteration is given by the following procedure.

- Initialize Q by

$$Q_i(x_i) \leftarrow \frac{1}{Z_i} \exp\{-\phi_u(x_i)\}.$$

- $\hat{Q}_i^{(m)}(l) \leftarrow \sum_{j\neq i} k^{(m)}(f_i, f_j)Q_j(l)$
- $\hat{Q}_i(x_i) \leftarrow \sum_{l'\in\mathcal{L}} \mu^{(m)}(x_i, l) \sum_m w^{(m)}\hat{Q}_i^{(m)}(l)$
- $Q_i(x_i) \leftarrow \exp\left\{-\psi_u(x_i) - \hat{Q}_i(x_i)\right\}$
- normalize $Q_i(x_i)$
- repeat until convergence.

2 End-to-end Learning and Inference of CRF

For the mean field approximation inference of a CRF, we can embed all the calculation into a CNN layer, and the iteration can be regarded as another RNN structure. Thus the inference step can be solely based on Neural Networks structure. For the training of this RNN-CRF layer, we use maximal log-likelihood strategy for the parameter updating. However, in this case the loss function of the Neural Networks will admite an intractable gradient.

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \log P(\mathbf{x}^{(n)}|\mathbf{I}^{(n)};\theta) = -\sum_{n=1}^{N}[E(\mathbf{x}|\mathbf{I}^{(n)};\theta) + \log Z(\mathbf{I}^{(n)};\theta)] \tag{2.1}$$

$$\Rightarrow \nabla_\theta \mathcal{L}(\theta) = -\sum_{n=1}^{N} \nabla_\theta[E(\mathbf{x}|\mathbf{I}^{(n)};\theta) + \log Z(\mathbf{I}^{(n)};\theta)]. \tag{2.2}$$

and a straightforward calculation leads to

$$\nabla_\theta \mathcal{L}(\theta) \log Z(\mathbf{I}^{(n)};\theta) = -\mathbb{E}_{\mathbf{x}^{(n)}\sim P(\mathbf{x}^{(n)}|\mathbf{I}^{(n)};\theta)} \nabla_\theta E(\mathbf{x}|\mathbf{I}^{(n)};\theta).$$

In oreder to address this issue, we use mean field approximation again for the calculation of the marginal expectation term in the gradient. Thus, we build an end-to-end CRF model.