Phrase Grounding by Soft-Label Chain Conditional Random Field – Supplementary Materials –

Jiacheng Liu **Julia Hockenmaier**
University of Illinois at Urbana-Champaign, Urbana, IL, USA 61801
{jl25, juliahmr}@illinois.edu

1 Correctness of the Modified Forward Algorithm for Soft-Label Chain CRFs

Theorem. *The modified forward algorithm computes the KL divergence loss for Soft-Label Chain CRFs.*

Proof. It is known that the iterations on forward variables α computes partition function $Z(\boldsymbol{x})$. We only need to prove that

$$G = \sum_t \sum_{y^t, y^{t-1}} q(y^t, y^{t-1}|\boldsymbol{x}) s(y^t, y^{t-1}, \boldsymbol{x}) - \sum_t \sum_{y^t} q(y^t|\boldsymbol{x}) \log q(y^t|\boldsymbol{x})$$

We first prove by induction that

$$g_{y^t}^t = \sum_{t'<t} \sum_{y^{t'}, y^{t'-1}} q(y^{t'}, y^{t'-1}|\boldsymbol{x}) s(y^{t'}, y^{t'-1}, \boldsymbol{x}) - \sum_{t'<t} \sum_{y^{t'}} q(y^{t'}|\boldsymbol{x}) \log q(y^{t'}|\boldsymbol{x}) + \varepsilon(y^t, \boldsymbol{x}) - \log q_{y^t}^t$$

Base Case: When $t = 0$,

$$g_{y^0}^0 = 0$$

Induction: By inductive hypothesis, we have

$$g_{y^{t-1}}^{t-1} = \sum_{t'<t-1} \sum_{y^{t'}, y^{t'-1}} q(y^{t'}, y^{t'-1}|\boldsymbol{x}) s(y^{t'}, y^{t'-1}, \boldsymbol{x}) - \sum_{t'<t-1} \sum_{y^{t'}} q(y^{t'}|\boldsymbol{x}) \log q(y^{t'}|\boldsymbol{x}) + \varepsilon(y^{t-1}, \boldsymbol{x}) - \log q_{y^{t-1}}^{t-1}$$

Therefore,

$$\begin{aligned}
g_{y^t}^t &= \sum_{y^{t-1}} [g_{y^{t-1}}^{t-1} + \tau(y^t, y^{t-1}, \boldsymbol{x})] q_{y^{t-1}}^{t-1} + \varepsilon(y^t, \boldsymbol{x}) - \log q_{y^t}^t \\
&= \sum_{t'<t-1} \sum_{y^{t'}, y^{t'-1}} q(y^{t'}, y^{t'-1}|\boldsymbol{x}) s(y^{t'}, y^{t'-1}, \boldsymbol{x}) - \sum_{t'<t-1} \sum_{y^{t'}} q(y^{t'}|\boldsymbol{x}) \log q(y^{t'}|\boldsymbol{x}) \\
&\quad + \sum_{y^{t-1}} [s(y^t, y^{t-1}, \boldsymbol{x}) - \log q_{y^{t-1}}^{t-1}] q_{y^{t-1}}^{t-1} + \varepsilon(y^t, \boldsymbol{x}) - \log q_{y^t}^t \\
&= \sum_{t'<t} \sum_{y^{t'}, y^{t'-1}} q(y^{t'}, y^{t'-1}|\boldsymbol{x}) s(y^{t'}, y^{t'-1}, \boldsymbol{x}) - \sum_{t'<t} \sum_{y^{t'}} q(y^{t'}|\boldsymbol{x}) \log q(y^{t'}|\boldsymbol{x}) + \varepsilon(y^t, \boldsymbol{x}) - \log q_{y^t}^t
\end{aligned}$$

By following a similar derivation in the induction step, we have

$$\begin{aligned}
G &= \sum_{y^T} g_{y^T}^T q_{y^T}^T \\
&= \sum_{t'<T} \sum_{y^{t'}, y^{t'-1}} q(y^{t'}, y^{t'-1}|\boldsymbol{x}) s(y^{t'}, y^{t'-1}, \boldsymbol{x}) - \sum_{t'<T} \sum_{y^{t'}} q(y^{t'}|\boldsymbol{x}) \log q(y^{t'}|\boldsymbol{x}) + \sum_{y^T} [s(y^T, y^{T-1}, \boldsymbol{x}) - \log q_{y^T}^T] q_{y^T}^T \\
&= \sum_t \sum_{y^t, y^{t-1}} q(y^t, y^{t-1}|\boldsymbol{x}) s(y^t, y^{t-1}, \boldsymbol{x}) - \sum_t \sum_{y^t} q(y^t|\boldsymbol{x}) \log q(y^t|\boldsymbol{x})
\end{aligned}$$

$\square$

Algorithm 1 Modified forward algorithm to compute the KL divergence loss for Soft-Label Chain CRFs

procedure $\text{SoftLabelChainCrfLoss}(\boldsymbol{q}, \varepsilon(y^t, \boldsymbol{x}), \tau(y^t, y^{t-1}, \boldsymbol{x}))$
  for all label y^0 **do**
    $\alpha^0_{y^0} \leftarrow 0$
    $g^0_{y^0} \leftarrow 0$
  for $t = 1 \dots T$ **do**
    for all label y^t **do**
      $\alpha^t_{y^t} \leftarrow \sum_{y^{t-1}} \left\{ \alpha^{t-1}_{y^{t-1}} \exp \left[\tau(y^t, y^{t-1}, \boldsymbol{x}) + \varepsilon(y^t, \boldsymbol{x}) \right] \right\}$
      $g^t_{y^t} \leftarrow \sum_{y^{t-1}} \left\{ \left[g^{t-1}_{y^{t-1}} + \tau(y^t, y^{t-1}, \boldsymbol{x}) \right] q^{t-1}_{y^{t-1}} + \left[\varepsilon(y^t, \boldsymbol{x}) - \log q^t_{y^t} \right] \right\}$
  $Z \leftarrow \sum_{y^T} \alpha^T_{y^T}$
  $G \leftarrow \sum_{y^T} g^T_{y^T} q^T_{y^T}$
  $L \leftarrow -G + \log Z$
  return L

2 Equivalence of Forward-Backward and Backpropagation in Soft-Label Chain CRFs

Lemma. *Given forward and backward variables and partition function computed by the forward-backward algorithm*

$$\alpha^t_{y^t} = \sum_{y^{t-1}} \alpha^{t-1}_{y^{t-1}} \exp s(y^t, y^{t-1}, \boldsymbol{x})$$
$$\beta^{t-1}_{y^{t-1}} = \sum_{y^t} \beta^t_{y^t} \exp s(y^t, y^{t-1}, \boldsymbol{x})$$
$$Z(\boldsymbol{x}) = \sum_{y^T} \alpha^T_{y^T}$$

We have

$$\frac{\partial Z(\boldsymbol{x})}{\partial \alpha^t_{y^t}} = \beta^t_{y^t}$$

Proof. By induction on t.
Base case: When $t = T$,

$$\frac{\partial Z(\boldsymbol{x})}{\partial \alpha^T_{y^T}} = 1 = \beta^T_{y^T}$$

Induction:

$$\begin{aligned} \frac{\partial Z(\boldsymbol{x})}{\partial \alpha^{t-1}_{y^{t-1}}} &= \sum_{y^t} \frac{\partial Z(\boldsymbol{x})}{\partial \alpha^t_{y^t}} \frac{\partial \alpha^t_{y^t}}{\partial \alpha^{t-1}_{y^{t-1}}} \\ &= \sum_{y^t} \beta^t_{y^t} \exp s(y^t, y^{t-1}, \boldsymbol{x}) \\ &= \beta^{t-1}_{y^{t-1}} \end{aligned}$$

□

Theorem. *Backpropagation on the loss of Soft-Label Chain CRFs gives gradient*

$$\frac{\partial L}{\partial s(y^t, y^{t-1}, \boldsymbol{x})} = -q(y^t, y^{t-1}|\boldsymbol{x}) + p(y^t, y^{t-1}|\boldsymbol{x})$$

where

$$p(y^t, y^{t-1}|\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x})} \alpha^{t-1}_{y^{t-1}} \beta^t_{y^t} \exp s(y^t, y^{t-1}, \boldsymbol{x})$$

Proof. By induction on t.
Base case: When $t = T$,

$$\begin{aligned} &\frac{\partial L}{\partial s(y^T, y^{T-1}, \boldsymbol{x})} \\ &= -q(y^T, y^{T-1}|\boldsymbol{x}) + \frac{1}{Z(\boldsymbol{x})} \frac{\partial Z(\boldsymbol{x})}{\partial s(y^T, y^{T-1}, \boldsymbol{x})} \\ &= -q(y^T, y^{T-1}|\boldsymbol{x}) + \frac{\alpha^{T-1}_{y^{T-1}} \exp s(y^T, y^{T-1}, \boldsymbol{x})}{Z(\boldsymbol{x})} \\ &= -q(y^T, y^{T-1}|\boldsymbol{x}) + p(y^T, y^{T-1}|\boldsymbol{x}) \end{aligned}$$

Induction:

$$\begin{aligned} &\frac{\partial L}{\partial s(y^t, y^{t-1}, \boldsymbol{x})} \\ &= -q(y^t, y^{t-1}|\boldsymbol{x}) + \frac{1}{Z(\boldsymbol{x})} \frac{\partial Z(\boldsymbol{x})}{\partial s(y^t, y^{t-1}, \boldsymbol{x})} \\ &= -q(y^t, y^{t-1}|\boldsymbol{x}) + \frac{1}{Z(\boldsymbol{x})} \frac{\partial Z(\boldsymbol{x})}{\partial \alpha^t_{y^t}} \frac{\partial \alpha^t_{y^t}}{\partial s(y^t, y^{t-1}, \boldsymbol{x})} \\ &= -q(y^t, y^{t-1}|\boldsymbol{x}) + \frac{\beta^t_{y^t} \alpha^{t-1}_{y^{t-1}} \exp s(y^t, y^{t-1}, \boldsymbol{x})}{Z(\boldsymbol{x})} \\ &= -q(y^t, y^{t-1}|\boldsymbol{x}) + p(y^t, y^{t-1}|\boldsymbol{x}) \end{aligned}$$

□