Video Event Recognition with Deep Hierarchical Context Model - Supplementary Material

Xiaoyang Wang and Qiang Ji
Dept. of ECSE, Rensselaer Polytechnic Institute, USA
{wangx16,jiq}@rpi.edu

1. Deep Context Model

1.1. Conditional Distributions of Different Units

Given the deep context model defined in the paper, we further provide the conditional distributions of different units given their adjacent units as:

$$P(h_{pi}=1|\mathbf{p},\mathbf{h}_r)=\sigma(\sum_j W^1_{ji}p_j/\sigma_{\mathbf{p}j}+\sum_k Q^1_{ik}h_{rk}+b_{\mathbf{h}_p i}) \tag{1}$$

$$P(h_{oj}=1|\mathbf{o},\mathbf{h}_r)=\sigma(\sum_i W^2_{ij}o_i/\sigma_{\mathbf{o}i}+\sum_k Q^2_{jk}h_{rk}+b_{\mathbf{h}_o j}) \tag{2}$$

$$P(h_{rk}=1|\mathbf{h}_p,\mathbf{h}_o,\mathbf{y})=\sigma(\sum_i Q^1_{ik}h_{pi}+\sum_j Q^2_{jk}h_{oj}+\sum_{k'} L_{kk'}y_{k'}+b_{\mathbf{h}_r k}) \tag{3}$$

$$P(y_k=1|\mathbf{e},\mathbf{c},\mathbf{h}_r,\mathbf{h}_s,\mathbf{y}_{-1})=\frac{\exp(\sum_f U^1_{fk}\tilde{e}_f+\sum_{f'}U^2_{f'k}\tilde{c}_{f'}+\sum_g L_{gk}h_{rg}+\sum_i T_{ik}h_{si}+\sum_j D_{jk}y_{-1,j}+b_{\mathbf{y}k})}{\sum_{k'}\exp(\sum_f U^1_{fk'}\tilde{e}_f+\sum_{f'}U^2_{f'k'}\tilde{c}_{f'}+\sum_g L_{gk'}h_{rg}+\sum_i T_{ik'}h_{si}+\sum_j D_{jk'}y_{-1,j}+b_{\mathbf{y}k'})} \tag{4}$$

$$P(y_{-1,j}=1|\mathbf{y},\mathbf{m}_{-1})=\frac{\exp(\sum_i F_{ij}m_{-1,i}+\sum_k D_{jk}y_k+b_{\mathbf{y}_{-1}j})}{\sum_{j'}\exp(\sum_i F_{ij'}m_{-1,i}+\sum_k D_{j'k}y_k+b_{\mathbf{y}_{-1}j'})} \tag{5}$$

$$P(m_{-1,j}=1|\mathbf{y}_{-1})=\frac{\exp(\sum_i F_{ji}y_{-1,i}+b_{\mathbf{m}_{-1}j})}{\sum_{j'}\exp(\sum_i F_{j'i}y_{-1,i}+b_{\mathbf{m}_{-1}j'})} \tag{6}$$

$$P(h_{si}=1|\mathbf{y},\mathbf{s})=\sigma(\sum_j G_{ji}s_j/\sigma_{\mathbf{s}j}+\sum_k T_{ik}y_k+b_{\mathbf{h}_s i}) \tag{7}$$

$$P(\mathbf{p}|\mathbf{h}_p)=\prod_i\frac{1}{\sigma_{\mathbf{p}i}\sqrt{2\pi}}\exp\left(-\frac{p_i-b_{\mathbf{p}i}-\sigma_{\mathbf{p}i}\sum_j W^1_{ij}h_{pj}}{2\sigma^2_{\mathbf{p}i}}\right) \tag{8}$$

$$P(\mathbf{o}|\mathbf{h}_o)=\prod_i\frac{1}{\sigma_{\mathbf{o}i}\sqrt{2\pi}}\exp\left(-\frac{o_i-b_{\mathbf{o}i}-\sigma_{\mathbf{o}i}\sum_j W^2_{ij}h_{oj}}{2\sigma^2_{\mathbf{o}i}}\right) \tag{9}$$

$$P(\mathbf{e}|\mathbf{y})=\prod_i\frac{1}{\sigma_{\mathbf{e}i}\sqrt{2\pi}}\exp\left(-\frac{e_i-b_{\mathbf{e}i}-\sigma_{\mathbf{e}i}\sum_j U^1_{ij}y_j}{2\sigma^2_{\mathbf{e}i}}\right) \tag{10}$$

$$P(\mathbf{c}|\mathbf{y})=\prod_i\frac{1}{\sigma_{\mathbf{c}i}\sqrt{2\pi}}\exp\left(-\frac{c_i-b_{\mathbf{c}i}-\sigma_{\mathbf{c}i}\sum_j U^2_{ij}y_j}{2\sigma^2_{\mathbf{c}i}}\right) \tag{11}$$

$$P(\mathbf{s}|\mathbf{h}_s)=\prod_i\frac{1}{\sigma_{\mathbf{s}i}\sqrt{2\pi}}\exp\left(-\frac{s_i-b_{\mathbf{s}i}-\sigma_{\mathbf{s}i}\sum_j G_{ij}h_{sj}}{2\sigma^2_{\mathbf{s}i}}\right) \tag{12}$$

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

where $\sigma(x) = 1/(1+\exp(-x))$ is the logistic function. In Equation 4, $y_k = 1$ would indicate the remaining units in $\mathbf{y}$ to be zero since $\mathbf{y}$ is defined through the 1-of-$\mathcal{K}$ coding scheme to indicate the class label for the current event sequence. The same rule is applied to Equation 5 and Equation 6. These conditional distributions are used in different phrase in the learning and inference of the model as discussed in the following.

1.2. Model Learning

Here, we given more details on the approximate learning of the proposed deep context model. Specifically, for estimating the data-dependent expectation, we replace the true posterior $P(\mathbf{h}|\mathbf{v};\theta)$ by the variational posterior $Q(\mathbf{h}|\mathbf{v};\boldsymbol{\mu})$. The mean-field approximation assumes all the hidden units are fully factorized as:

$$Q(\mathbf{h}|\mathbf{v};\boldsymbol{\mu}) = \left(\prod_i q(h_{pi}|\mathbf{v})\right)\left(\prod_j q(h_{oj}|\mathbf{v})\right)\left(\prod_k q(h_{rk}|\mathbf{v})\right)\left(\prod_g q(h_{sg}|\mathbf{v})\right) \tag{13}$$

where $\boldsymbol{\mu} = \{\boldsymbol{\mu}_p, \boldsymbol{\mu}_o, \boldsymbol{\mu}_r, \boldsymbol{\mu}_s\}$ are the mean field variational parameters with $q(h_i = 1) = \mu_i$. The estimation then proceeds by finding the parameters $\boldsymbol{\mu}$ that maximizes the variational lower bound of the log conditional likelihood for fixed θ, which results in iteratively updating $\boldsymbol{\mu}$ for different hidden units through the mean-field fixed point equations.

For estimating the model's expectation, we use the MCMC based stochastic approximation procedure. It first randomly initialize M Markov chains with samples of $\{\mathbf{y}^{0,j}, \mathbf{h}_r^{0,j}, \mathbf{h}_p^{0,j}, \mathbf{h}_o^{0,j}, \mathbf{p}^{0,j}, \mathbf{o}^{0,j}, \mathbf{e}^{0,j}, \mathbf{c}^{0,j}, \mathbf{y}_{-1}^{0,j}, \mathbf{m}_{-1}^{0,j}, \mathbf{h}_s^{0,j}, \mathbf{s}^{0,j}\}_{j=1}^{M}$. For each Markov chain j from 1 to M, the $(t+1)$th step samples $\mathbf{y}^{t+1,j}$, $\mathbf{h}_r^{t+1,j}$, $\mathbf{h}_p^{t+1,j}$, $\mathbf{h}_o^{t+1,j}$, $\mathbf{p}^{t+1,j}$, $\mathbf{o}^{t+1,j}$, $\mathbf{e}^{t+1,j}$, $\mathbf{c}^{t+1,j}$, $\mathbf{y}_{-1}^{t+1,j}$, $\mathbf{m}_{-1}^{t+1,j}$, $\mathbf{h}_s^{t+1,j}$, $\mathbf{s}^{t+1,j}$ given the samples from the tth step as $\mathbf{y}^{t,j}$, $\mathbf{h}_r^{t,j}$, $\mathbf{h}_p^{t,j}$, $\mathbf{h}_o^{t,j}$, $\mathbf{p}^{t,j}$, $\mathbf{o}^{t,j}$, $\mathbf{e}^{t,j}$, $\mathbf{c}^{t,j}$, $\mathbf{y}_{-1}^{t,j}$, $\mathbf{m}_{-1}^{t,j}$, $\mathbf{h}_s^{t,j}$, $\mathbf{s}^{t,j}$ by running a Gibbs sampler with conditional distributions given in Equations 1 to 12. The M sampled Markov particles are then used to estimate the model's expectation in the corresponding step of the model optimization.

1.3. Model Inference

The detailed algorithm for inferring the probability $P(y_k = 1|\mathbf{e}, \mathbf{c}, \mathbf{p}, \mathbf{o}, \mathbf{s}, \mathbf{m}_{-1}; \theta)$ through Gibbs sampling is summarized in Algorithm 1.

Algorithm 1: Inference of $P(y|\mathbf{e}, \mathbf{c}, \mathbf{p}, \mathbf{o}, \mathbf{s}, \mathbf{m}_{-1})$ with Gibbs sampling.

Data: the input observation vectors $\mathbf{e}$, $\mathbf{c}$, $\mathbf{p}$, $\mathbf{o}$, $\mathbf{s}$, and $\mathbf{m}_{-1}$ for the query event sequence; model parameter set θ
Result: $P(y_k = 1|\mathbf{e}, \mathbf{c}, \mathbf{p}, \mathbf{o}, \mathbf{s}, \mathbf{m}_{-1}; \theta)$ for $k = 1, \ldots, K$
for $chain = 1 \to C$ **do**
  Randomly initialize $\mathbf{h}_p^0$, $\mathbf{h}_o^0$ and $\mathbf{y}^0$;
  for $t = 0 \to T$ **do**
    Sample $\mathbf{h}_r^t$ given $\mathbf{h}_p^t$, $\mathbf{h}_o^t$, and $\mathbf{y}^t$ with Equation 3;
    Sample $\mathbf{h}_p^{t+1}$ given $\mathbf{h}_r^t$ and $\mathbf{p}$ with Equation 1;
    Sample $\mathbf{h}_o^{t+1}$ given $\mathbf{h}_r^t$ and $\mathbf{o}$ with Equation 2;
    Sample $\mathbf{h}_s^{t+1}$ given $\mathbf{y}^t$ and $\mathbf{s}$ with Equation 7;
    Sample $\mathbf{y}_{-1}^{t+1}$ given $\mathbf{y}^t$ and $\mathbf{m}_{-1}$ with Equation 5;
    Sample $\mathbf{y}^{t+1}$ given $\mathbf{h}_r^t$, $\mathbf{h}_s^{t+1}$, $\mathbf{y}_{-1}^{t+1}$, $\mathbf{e}$ and $\mathbf{c}$ with Equation 4;
  end
end
Collect the last T' samples of $\mathbf{y}$ from each chain;
Calculate $P(y|\mathbf{e}, \mathbf{c}, \mathbf{p}, \mathbf{o}, \mathbf{s}, \mathbf{m}_{-1})$ with the collected samples;