DISCRIMINATIVE PROBABILISTIC FRAMEWORK FOR GENERALIZED MULTI-INSTANCE LEARNING

Anh T. Pham, Raviv Raich, Xiaoli Z. Fern, Weng-Keen Wong, and Xinze Guan

School of Electrical Engineering and Computer Science
Oregon State University, Corvallis, Oregon 97331-5501
{phaman,raich,xfern,wong}@eecs.oregonstate.edu, guanxinze@hotmail.com

1. Compute α. Recall that $\alpha_i(k) \triangleq p(n_{bi} = k|\mathbf{X}_b, \mathbf{w}')$. Consequently, $\alpha_{i+1}(k)$ can be computed by marginalizing $p(n_{b(i+1)} = k, n_{bi} = l, y_{b(i+1)} = c|\mathbf{X}_b, \mathbf{w}')$ over $y_{b(i+1)}$ and n_{bi} as follows

$$\begin{aligned}&\alpha_{i+1}(k) = p(n_{b(i+1)} = k|\mathbf{X}_b, \mathbf{w}')\\ &= \sum_{l=0}^{n_b}\sum_{c=0}^{1} p(n_{b(i+1)} = k, n_{bi} = l, y_{b(i+1)} = c|\mathbf{X}_b, \mathbf{w}')\end{aligned}$$

Using conditional rule and the independent assumption among instance labels we obtain

$$\begin{aligned}\alpha_{i+1}(k) = &\sum_{l=0}^{n_b}\sum_{c=0}^{1} p(n_{b(i+1)} = k|n_{bi} = l, y_{b(i+1)} = c)\\ &p(n_{bi} = l|\mathbf{X}_b, \mathbf{w}')p(y_{b(i+1)} = c|\mathbf{x}_{b(i+1)}, \mathbf{w}')\end{aligned}$$

Replace $p(n_{b(i+1)} = k|n_{bi} = l, y_{b(i+1)} = c) = \mathbb{I}[k = l + c]$ we obtain the result for $\alpha_{i+1}(k)$ as follows

$$\begin{aligned}\alpha_{i+1}(k) = &p(y_{b(i+1)} = 1|\mathbf{x}_{b(i+1)}, \mathbf{w}')\alpha_i(k-1)\\ &+ p(y_{b(i+1)} = 0|\mathbf{x}_{b(i+1)}, \mathbf{w}')\alpha_i(k). \qquad (1)\end{aligned}$$

The intuition behind above equation is that: there are two cases that there are k positive instances in the first $(i+1)$ instances. First, the $(i+1)$th instance is positive and there are $k-1$ positive instances in the first i instances. Second, the $(i+1)$th instance is negative and there are k positive instances in the first i instances.

2. Compute β. Recall that $\beta_i(k) \triangleq p(Y_b|n_{bi} = k, \mathbf{X}_b, \mathbf{w}')$. Consequently, $\beta_i(k)$ can be computed by marginalizing $p(Y_b|n_{bi} = k, \mathbf{X}_b, \mathbf{w}')$ over $y_{b(i+1)}$ and $n_{b(i+1)}$ as follows

$$\begin{aligned}&\beta_i(k) = p(Y_b|n_{bi} = k, \mathbf{X}_b, \mathbf{w}')\\ &= \sum_{l=0}^{n_b}\sum_{c=0}^{1} p(Y_b, n_{b(i+1)} = l, y_{b(i+1)} = c|n_{bi} = k, \mathbf{X}_b, \mathbf{w}')\end{aligned}$$

Using conditional rule and the independent assumption among instance labels we obtain

$$\begin{aligned}&\beta_i(k) = \sum_{l=0}^{n_b}\sum_{c=0}^{1} p(Y_b|n_{b(i+1)} = l, \mathbf{X}_b, \mathbf{w}')\\ &p(n_{b(i+1)} = l|y_{b(i+1)} = c, n_{bi} = k)p(y_{b(i+1)} = c|\mathbf{x}_{b(i+1)}, \mathbf{w}')\end{aligned}$$

This work is partially supported by the National Science Foundation grants CCF-1254218, DBI-1356792, and IIS-1055113.

Replace $p(n_{b(i+1)} = l|y_{b(i+1)} = c, n_{bi} = k) = 1$ if $l = k + c$ we obtain the result for $\beta_i(k)$ as follows

$$\begin{aligned}\beta_i(k) = &p(y_{b(i+1)} = 1|\mathbf{x}_{b(i+1)}, \mathbf{w}')\beta_{i+1}(k+1)\\ &+ p(y_{b(i+1)} = 0|\mathbf{x}_{b(i+1)}, \mathbf{w}')\beta_{i+1}(k). \qquad (2)\end{aligned}$$

3. Compute $p(y_{bi} = 1|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}')$. Using conditional probability we obtain

$$\begin{aligned}&p(y_{bi} = 1|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}') \qquad (3)\\ &= \frac{p(y_{bi} = 1, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}')}{p(y_{bi} = 0, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}') + p(y_{bi} = 1, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}')}.\end{aligned}$$

First, we obtain $p(y_{bi} = 1, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}')$ by marginalizing over n_{bi} and $n_{b(i-1)}$ as follows

$$\begin{aligned}&p(y_{bi} = 1, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}') =\\ &\sum_{l=0}^{n_b}\sum_{k=0}^{n_b} p(y_{bi} = 1, Y_b, n_{bi} = l, n_{b(i-1)} = k|\mathbf{X}_b, \mathbf{w}', \mathbf{v}').\end{aligned}$$

Using the conditional rule and the independent assumption among instance labels we obtain

$$\begin{aligned}&p(y_{bi} = 1, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}') =\\ &\sum_{l=0}^{n_b}\sum_{k=0}^{n_b} p(Y_b|n_{bi} = l, \mathbf{X}_b, \mathbf{w}')p(n_{bi} = l|n_{b(i-1)} = k, y_{bi} = 1)\\ &\qquad p(n_{b(i-1)} = k|\mathbf{X}_b, \mathbf{w}')p(y_{bi} = 1|\mathbf{x}_{bi}, \mathbf{w}').\end{aligned}$$

Using the fact that $p(n_{bi} = l|n_{b(i-1)} = k, y_{bi} = 1) = \mathbb{I}[l = k+1]$, we obtain

$$\begin{aligned}&p(y_{bi} = 1, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}') =\\ &\sum_{k=0}^{n_b} p(Y_b|n_{bi} = k+1, \mathbf{X}_b, \mathbf{w}')\times\\ &\qquad p(n_{b(i-1)} = k|\mathbf{X}_b, \mathbf{w}')p(y_{bi} = 1|\mathbf{x}_{bi}, \mathbf{w}').\end{aligned}$$

Using the definition of α and β we obtain

$$\begin{aligned}&p(y_{bi} = 1, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}') =\\ &\sum_{k=0}^{n_b} \beta_i(k+1)\alpha_{i-1}(k)p(y_{bi} = 1|\mathbf{x}_{bi}, \mathbf{w}'). \qquad (4)\end{aligned}$$

Using similar proof technique, we obtain

$$\begin{aligned}&p(y_{bi} = 0, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}') =\\ &\sum_{k=0}^{n_b} \beta_i(k)\alpha_{i-1}(k)p(y_{bi} = 0|\mathbf{x}_{bi}, \mathbf{w}'). \qquad (5)\end{aligned}$$

Then, we obtain the instance membership probability $p(y_{bi} = 1|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}')$ as follows

$$p(y_{bi} = 1|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}') = \frac{\pi_1}{\pi_1 + \pi_0} \tag{6}$$

where

$$\pi_1 = \sum_{k=0}^{n_b} \alpha_{i-1}(k)\beta_i(k+1)p(y_{bi} = 1|\mathbf{x}_{bi}, \mathbf{w}'),$$
$$\pi_0 = \sum_{k=0}^{n_b} \alpha_{i-1}(k)\beta_i(k)p(y_{bi} = 0|\mathbf{x}_{bi}, \mathbf{w}').$$

4. Compute $p(n_{bi} = n|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}')$. Using conditional probability we obtain

$$p(n_{bi} = n|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}') = \frac{p(n_{bi} = n, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}')}{\sum_{k=0}^{n_b} p(n_{bi} = k, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}')},$$

where $p(n_{bi} = k, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}')$ is computed by using the conditional rule

$$\begin{aligned} p(n_{bi} = k, Y_b|\mathbf{X}_b, \mathbf{w}', \mathbf{v}') =& p(Y_b|n_{bi} = k, \mathbf{v}') \\ & \times p(n_{bi} = k|\mathbf{X}_b, \mathbf{w}'). \end{aligned}$$

Using the definition of α and β we obtain the probability of the number of positive instances $p(n_{bi} = n|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}')$ as follows

$$p(n_{bi} = n|Y_b, \mathbf{X}_b, \mathbf{w}', \mathbf{v}') = \frac{\alpha_i(n)\beta_i(n)}{\sum_{k=0}^{n_b} \alpha_i(k)\beta_i(k)}. \tag{7}$$