Supplementary Material for "Conditional Random Fields with High-Order Features for Sequence Labeling"

Nan Ye **Wee Sun Lee**
Department of Computer Science
National University of Singapore
{yenan,leews}@comp.nus.edu.sg

Hai Leong Chieu
DSO National Laboratories
chaileon@dso.org.sg

Dan Wu
Singapore MIT Alliance
National University of Singapore
dwu@nus.edu.sg

An example illustrating the computation of the forward variable and marginal

Assume $\mathbf{x}$ is "Peter goes to Britain and France annually.". Assume there are 9 binary features defined by boolean predicates as in Table 1, and each $\lambda_i = 1$. The label set is $\{P, O, L\}$ where P represents *Person*, L represents *Location* and O represents *Others*.

i	$f_i(\mathbf{x}, \mathbf{y}, j)$
1	$x_j = Peter \wedge y_j = P$
2	$x_j = goes \wedge y_j = O$
3	$x_j = to \wedge y_j = O$
4	$x_j = Britain \wedge y_j = L$
5	$x_j = and \wedge y_j = O$
6	$x_j = France \wedge y_j = L$
7	$x_j = annually \wedge y_j = O$
8	$x_j = . \wedge y_j = O$
9	$y_{j-2}y_{j-1}y_j = LOL$

Table 1: List of features.

Table 2 shows the sum of the weights for features with the same label pattern at each position.

$t\backslash\mathbf{z}$	P	O	L	LOL
0	0	0	0	0
1	1	0	0	0
2	0	1	0	0
3	0	1	0	0
4	0	0	1	0
5	0	1	0	0
6	0	0	1	1
7	0	1	0	0
8	0	1	0	0

Table 2: $\sum_{i:\mathbf{z}^i=\mathbf{z}} \lambda_i g_i(x, t)$

We have $\mathcal{P} = \{\epsilon, P, O, L, LO\}$, and $\mathcal{S} = \{\epsilon P, O, L, OL\}$. The tables for $\ln \alpha_{\mathbf{x}}$ and $\ln \beta_{\mathbf{x}}$ are shown in Table 3 and Table 4 respectively.

$t\backslash\mathbf{p}^i$	ϵ	P	O	L	LO
0	0.00	-∞	-∞	-∞	-∞
1	-∞	1.00	0.00	0.00	-∞
2	-∞	1.55	2.31	1.55	1.00
3	-∞	3.10	3.87	3.12	2.55
4	-∞	4.65	4.42	5.65	3.10
5	-∞	6.21	6.35	6.21	6.65
6	-∞	7.76	7.52	9.21	6.21
7	-∞	9.60	9.45	9.59	10.21
8	-∞	11.14	11.91	11.14	10.59

Table 3: $\ln \alpha_{\mathbf{x}}(t, \mathbf{p}^i)$

$t\backslash\mathbf{s}^i$	ϵ	P	O	L	OL
1	-∞	12.14	10.91	11.14	9.59
2	-∞	9.59	10.36	9.59	9.04
3	-∞	8.04	7.90	8.04	8.66
4	-∞	6.21	5.97	7.66	4.65
5	-∞	4.65	4.80	4.65	5.10
6	-∞	3.10	2.87	4.10	1.55
7	-∞	1.55	2.31	1.55	1.00
8	-∞	0.00	1.00	0.00	-∞
9	0.00	-∞	-∞	-∞	-∞

Table 4: $\ln \beta_{\mathbf{x}}(t, \mathbf{s}^i)$

We illustrate the computation of $\alpha_{\mathbf{x}}$ with $\alpha_{\mathbf{x}}(6, L)$. The condition $(\mathbf{p}^i, y) : \mathbf{p}^k \leq^s_{\mathcal{P}} \mathbf{p}^i y$ with $\mathbf{p}^k = L$ gives us the following 5 label sequences as $(\mathbf{p}^i, y)$: $\{\epsilon L, PL, OL, LL, LOL\}$.

$$\begin{aligned}
& \alpha_{\mathbf{x}}(6, L) \\
= \ & \alpha_{\mathbf{x}}(5, \epsilon)\Psi^p_{\mathbf{x}}(6, L) + \alpha_{\mathbf{x}}(5, P)\Psi^p_{\mathbf{x}}(6, PL) + \alpha_{\mathbf{x}}(5, O)\Psi^p_{\mathbf{x}}(6, OL) + \\
& \alpha_{\mathbf{x}}(5, L)\Psi^p_{\mathbf{x}}(6, LL) + \alpha_{\mathbf{x}}(5, LO)\Psi^p_{\mathbf{x}}(6, LOL) \\
= \ & 0\Psi^p_{\mathbf{x}}(6, L) + \alpha_{\mathbf{x}}(5, P)e + \alpha_{\mathbf{x}}(5, O)e + \alpha_{\mathbf{x}}(5, L)e + \alpha_{\mathbf{x}}(5, LO)e^2.
\end{aligned}$$

We also have

$$\begin{aligned}
Z_{\mathbf{x}} &= \alpha_{\mathbf{x}}(8, \epsilon) + \alpha_{\mathbf{x}}(8, P) + \alpha_{\mathbf{x}}(8, O) + \alpha_{\mathbf{x}}(8, L) + \alpha_{\mathbf{x}}(8, LO) \\
&= \beta_{\mathbf{x}}(1, \epsilon) + \beta_{\mathbf{x}}(1, P) + \beta_{\mathbf{x}}(1, O) + \beta_{\mathbf{x}}(1, L) + \beta_{\mathbf{x}}(1, LO) \\
&= e^{12.696}.
\end{aligned}$$

The table for $P(\mathbf{y}_{t-|\mathbf{z}^i|+1:t} = \mathbf{z}^i | \mathbf{x})$ is shown in Table 5.

$t\backslash\mathbf{z}^i$	P	O	L	LOL
1	0.576	0.212	0.212	0.000
2	0.212	0.576	0.212	0.000
3	0.212	0.576	0.212	0.026
4	0.160	0.160	0.680	0.083
5	0.160	0.681	0.160	0.007
6	0.160	0.160	0.681	0.391
7	0.212	0.576	0.212	0.007
8	0.212	0.576	0.212	0.083

Table 5: $P(\mathbf{y}_{t-|\mathbf{z}^i|+1:t} = \mathbf{z}^i | \mathbf{x})$

We illustrate the computation of $P(\mathbf{y}_{4:6} = LOL|\mathbf{x})$ from the forward and backward variables. Observe that there is only one way to satisfy the condition $(i, j) : \mathbf{z}_{1:|\mathbf{z}|-1} \leq^s \mathbf{p}^i, \mathbf{z}_{2:|\mathbf{z}|} \leq^p \mathbf{s}^j$ with $\mathbf{z} = LOL$. Hence,

$$\begin{aligned} P(\mathbf{y}_{4:6} = LOL|\mathbf{x}) &= \frac{\alpha_{bx}(5, LO)\beta_{\mathbf{x}}(5, OL)O_x(6, LO, OL, LOL)}{Z_{\mathbf{x}}W_{\mathbf{x}}} \\ &= \frac{\alpha_{\mathbf{x}}(5, LO)\beta_{\mathbf{x}}(5, OL)e^1}{Z_{\mathbf{x}}e^1}. \end{aligned}$$