Input: a training set $\{(\mathbf{s}_i, y_i)|i = 1...n\}$, where i is an index corresponding to a particular sentence pair $\mathbf{s}_i$, and y_i is the training label.

1: **initialize** parameter vector $\theta \leftarrow \mathbf{0}$
2: **for** $i \leftarrow 1$ **to** n **do**
3: **extract** all possible word pairs $\mathbf{w}_i = w_1, w_2, ..., w_m$ and their features from the sentence pair $\mathbf{s}_i$
4: **end for**
5: **for** $l \leftarrow 1$ **to** maximum iterations **do**
6: **for** $i \leftarrow 1$ **to** n **do**
7: $(y'_i, \mathbf{z}'_i) \leftarrow \arg\max_{y_i, \mathbf{z}_i} P(\mathbf{z}_i, y_i|\mathbf{w}_i; \theta)$
8: **if** $y'_i \neq y_i$ **then**
9: $\mathbf{z}^*_i \leftarrow \arg\max_{\mathbf{z}_i} P(\mathbf{z}_i|\mathbf{w}_i, y_i; \theta)$
10: $\theta \leftarrow \theta + f(\mathbf{z}^*_i, \mathbf{w}_i) - f(\mathbf{z}'_i, \mathbf{w}_i)$
11: **end if**
12: **end for**
13: **end for**
14: **return** model parameters θ