An Extraordinarily Brief Description of a Hybrid Model for Co-reference Resolution

Jason D. M. Rennie
jrennie@csail.mit.edu

November 4, 2004

Let x_i represent the i^{th} noun phrase (NP). Let A be the set of indices of proper NPs; let B be the set of indices of non-proper NPs. Let y_i represent the entity to which x_i refers. We define a pairwise indicator variable,

$$y_{ij} = \begin{cases} 1 & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases} . \tag{1}$$

Our model involves two parts, a model on proper NPs, which is identical to that of (McCallum & Wellner, 2003), and a model on non-proper NPs. Let y_A denote the proper NP labels; let y_B denote the non-proper NP labels. From a high level, we can see the model in terms of Bayes' Law:

$$P(\vec{y}|\vec{x}) = P(y_A|\vec{x})P(y_B|y_A, \vec{x}) \tag{2}$$

The first part, $P(y_A|\vec{x})$, is simply a product of pairwise potentials (normalized over possible configurations of the labels—note that the normalization constant is identical to the one used by (McCallum & Wellner, 2003)). We further decompose the second part, $P(y_B|y_A, \vec{x})$. Let $B_i = \{j \in B | j < i\}$. Then the conditional probability of the label of some non-proper NP is

$$P(y_i|y_A, y_{B_i}, \vec{x}) = \frac{\sum_{j<i|y_{ij}=1} \psi(x_i, x_j, 1)}{\sum_{j<i} \psi(x_i, x_j, 1)} \qquad \forall i \in B. \tag{3}$$

The second part of the model, $P(y_B|y_A, \vec{x})$, is a product of these conditionals. Hence, our joint model is

$$P(\vec{y}|\vec{x}) = \frac{\prod_{i,j\in A} \psi(x_i, x_j, y_{ij})}{\sum_{y_A} \prod_{i,j\in A} \psi(x_i, x_j, y_{ij})} \prod_{i\in B} \frac{\sum_{j<i|y_{ij}=1} \psi(x_i, x_j, 1)}{\sum_{j<i} \psi(x_i, x_j, 1)}. \tag{4}$$

References

McCallum, A., & Wellner, B. (2003). Toward conditional models of identity uncertainty with application to proper noun coreference. *Proceedings of the IJCAI Workshop on Information Integration on the Web.*