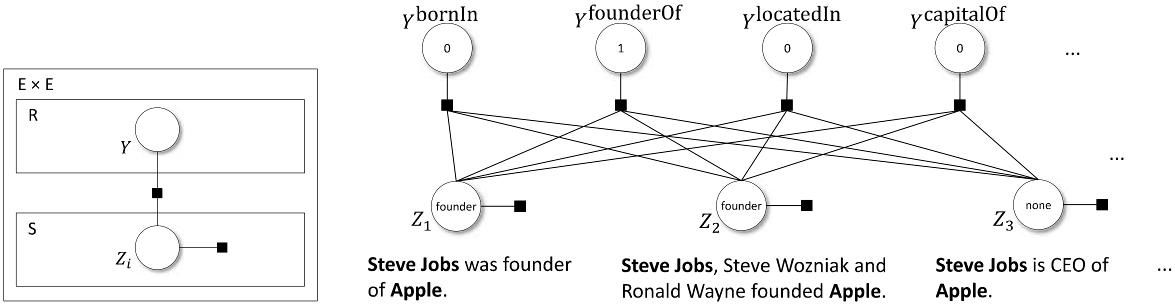

(a) (b)

Figure 1: (a) Network structure depicted as plate model and (b) an example network instantiation for the pair of entities `Steve Jobs`, `Apple`.

importantly, also the distinct value `none`. Z_i should be assigned a value $r \in R$ only when x_i expresses the ground fact $r(\mathbf{e})$, thereby modeling sentence-level extraction.

Figure 1(b) shows an example instantiation of the model with four relation names and three sentences.

3.2 A Joint, Conditional Extraction Model

We use a conditional probability model that defines a joint distribution over all of the extraction random variables defined above. The model is undirected and includes repeated factors for making sentence level predictions as well as globals factors for aggregating these choices.

For each entity pair $\mathbf{e} = (e_1, e_2)$, define $\mathbf{x}$ to be a vector concatenating the individual sentences $x_i \in S_{(e_1,e_2)}$, $\mathbf{Y}$ to be vector of binary Y^r random variables, one for each $r \in R$, and $\mathbf{Z}$ to be the vector of Z_i variables, one for each sentence x_i. Our conditional extraction model is defined as follows:

$$p(\mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z} | \mathbf{x}; \theta) \stackrel{\text{def}}{=} \frac{1}{Z_x} \prod_r \Phi^{\text{join}}(y^r, \mathbf{z}) \prod_i \Phi^{\text{extract}}(z_i, x_i)$$

where the parameter vector θ is used, below, to define the factor Φ^{extract}.

The factors Φ^{join} are deterministic OR operators

$$\Phi^{\text{join}}(y^r, \mathbf{z}) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } y^r = true \wedge \exists i : z_i = r \\ 0 & \text{otherwise} \end{cases}$$

which are included to ensure that the ground fact $r(\mathbf{e})$ is predicted at the aggregate level for the assignment $Y^r = y^r$ only if at least one of the sentence level assignments $Z_i = z_i$ signals a mention of $r(\mathbf{e})$.

The extraction factors Φ^{extract} are given by

$$\Phi^{\text{extract}}(z_i, x_i) \stackrel{\text{def}}{=} \exp\left(\sum_j \theta_j \phi_j(z_i, x_i)\right)$$

where the features ϕ_j are sensitive to the relation name assigned to extraction variable z_i, if any, and cues from the sentence x_i. We will make use of the Mintz *et al.* (2009) sentence-level features in the expeiments, as described in Section 7.

3.3 Discussion

This model was designed to provide a joint approach where extraction decisions are almost entirely driven by sentence-level reasoning. However, defining the Y^r random variables and tying them to the sentence-level variables, Z_i, provides a direct method for modeling weak supervision. We can simply train the model so that the Y variables match the facts in the database, treating the Z_i as hidden variables that can take any value, as long as they produce the correct aggregate predictions.

This approach is related to the multi-instance learning approach of Riedel *et al.* (2010), in that both models include sentence-level and aggregate random variables. However, their sentence level variables are binary and they only have a single aggregate variable that takes values $r \in R \cup \{$`none`$\}$, thereby ruling out overlapping relations. Additionally, their aggregate decisions make use of Mintz-style aggregate features (Mintz et al., 2009), that collect evidence from multiple sentences, while we use