semantic vector $V(w)$. The similarity between $R_\alpha(w_1)$ and $R_\alpha(w_2)$ is represented as the cosine of the angle between $V(w_1)$ and $V(w_2)$. Two special values are assigned to two exceptional cases: (i) when no relationship R_α is decoded in both contexts; (ii) when the relationship R_α is decoded only for one context.

In matching parsing relationships in a context pair, if only exact node match counts, very few cases can be covered, hence significantly reducing the effect of the parser in this task. To solve this problem, LSA is used as a type of synonym expansion in matching. For example, using LSA, the following word similarity values are generated:

similarity(good, good)	1.00
similarity(good, pretty)	0.79
similarity(good, great)	0.72
……	

Given a context pair of a noun keyword, suppose the first context involves a relationship *has-adjective-modifier* whose value is *good,* and the second context involves the same relationship *has-adjective-modifier* with the value *pretty,* then the system assigns 0.79 as the similarity value for this relationship pair.

To facilitate the maximum entropy modeling in the later stage, all the three categories of the resulting similarity values are discretized into 10 integers. Now the pairwise context similarity is represented as a set of similarity features, e.g.

{VSM-Trigger-Words-Similairty-equal-to-2,
LSA-Trigger-Words-Similarity-equal-to-1,
LSA-Subject-Similarity-equal-to-2}.

In addition to the three categories of basic context similarity features defined above, we also define induced context similarity features by combining basic context similarity features using the logical *and* operator. With induced features, the context similarity vector in the previous example is represented as

{VSM-Trigger-Word-Similairty-equal-to-2,
LSA- Trigger-Word-Similarity-equal-to-1,
LSA-Subject-Similarity-equal-to-2,
[VSM-Similairty-equal-to-2 *and*
LSA-Trigger-Word-Similarity-equal-to-1],
[VSM-Similairty-equal-to-2 *and*
LSA-Subject-Similarity-equal-to-2],
………
[VSM-Trigger-Word-Similairty-equal-to-2
and LSA-Trigger-Word-Similarity-equal-to-1
and LSA-Subject-Similarity-equal-to-2]
}

The induced features provide direct and fine-grained information, but suffer from less sampling space. Combining basic features and induced features under a smoothing scheme, maximum entropy modeling may achieve optimal performance.

Using the context similarity features defined above, the training corpora for the context pair classification model is in the following format:

Instance_0 tag="positive" {VSM-Trigger-Word-Similairty-equal-to-2, …}
Instance_1 tag="negative" {VSM-Trigger-Word-Similairty-equal-to-0, …}
…………..
where *positive* tag denotes a context pair associated with same sense, and *negative* tag denotes a context pair associated with different senses.

The maximum entropy modeling is used to compute the conditional probabilities $\Pr(S_i = S_j | CS_{i,j})$ and $\Pr(S_i \neq S_j | CS_{i,j})$: once the context pair $CS_{i,j}$ is represented as $\{f_\alpha\}$, the conditional probability is given as

$$\Pr(t|\{f_\alpha\}) = \frac{1}{Z} \prod_{f \in \{f_\alpha\}} w_{t,f} \qquad (1)$$

where $t \in \{S_i = S_j, S_i \neq S_j\}$, Z is the normalization factor, $w_{t,f}$ is the weight associated with tag t and feature f. Using the training corpora constructed above, the weights can be computed based on Iterative Scaling algorithm (Pietra etc. 1995) The exponential prior smoothing scheme (Goodman 2003) is adopted in the training.