Ling 572 Reading 3

Daniel Campos `dacampos@uw.edu`

02/28/2019

1 Q1: What does training data look like? That is, a classifier is trained with (x, y) pairs. For this reranking problem,what is x and what is y?

The traning data for parsing x is the input sentence and y is the target parse. In the reranking problem, the goal is to rank all possible parse trees by how likley they are to occour for the given sentence.

2 Q2: What happens at the test time? That is, what formula(s) one needs to calculate in order to determine the correct ranking of the candidate parse trees?

At test time a list of all possible parse tree with their plausability scores are produced. Then, to produce the predicted parse you select the parse that has the highest plasability which is produced by $w \cdot h(x)$ where w is learned weight vector and $h(x)$ is the feature vector represenation.

3 Q3:Conceptually, a parse tree is representedas a feature vector. What are the features?What are the feature values? How many features are there?

Each feature is a possible tree fragment. The feature values are the amount of times it appears in the full parse tree or the given example. The total features is all of possible tree fragments that occour in the entire training data.

4 Q4: In practice, is it necessary to represent a parse tree as a feature vector? Why or why not?

Yes it is necessary since there must be a vector represenation that can be multiplied by the learned weight matrix to produce a plasabulity score. If a feature vector did not exist it would not be possible to rank parses.