Reading Assignment 4

Q1. Training data will consist of input/output pairs {s_i, t_i} where each s_i is a sentence and each t_i is the correct tree for that sentence, and feature vector associated to each tree. The testing data will consists of input sentences along with a set of candidates associated to each sentence. Each candidate is denoted by x_{ij}, where *j*'th refers to the *i*'th sentence in the training data and each candidate is also represented by a feature vector $h(x_{ij})$.

Q2. The model file will store the ranking score for each candidate taken from the training data, along with a feature vector associated to each candidate. The ranking score is calculated from argmax of $w \cdot h(x)$ where *w* is the optimal weight and $h(x)$ is the feature vector representing each candidate.

Q3. Formula during test time will be $F(x) = \sum_{(i,\ j)} \alpha_{i,\ j} (h(x_{i1}) \cdot h(x) - h(x_{ij}) \cdot h(x))$. For each sentence, we will choose the F(x) that ranks the correct parse above all other candidate sentences, which is also argmax $F(x)$ among all candidate sentences.

Q4. Each feature corresponds to a tree fragment with more than one node, along with the count of occurrence of that tree fragment. The number of features will be the total count of unique tree fragments with more than one node found in the tree.

Q5. It is necessary to break the trees into smaller fragments and then group them in the form of features. We would construct appropriate kernels over these features and eventually combine these kernels up for the entire tree. We can then focus only on the "atomic" parts of a highly structured object such as a tree, and then extends our calculation recursively without losing the structure of the object.

Submitted by:
Wee Teck Tan