11-791: Homework 2

Instructed by *Dr. Eric Nyberg*

Due on Sep 23, 2013

Troy, Zhenhao Hua zhua

1 Introduction

The aggregate AE combines five annotators sequentially:

1. QADescriptor
   This part extracts question and answer body from the document
2. TokenDescriptor
   Using the output of the previous step, this engine annotates individual tokens for each answer and question. For the efficiency of scoring function, this step will each asnwer and question with its own tokens in an FSArray feature
3. NGramDescriptor
   Similar to TokenDescriptor, this step annotates all the 1-3 grams in each answer and question.
4. ScoreDescriptor
   This is the scoring component of the task. Using the token and NGram information from previous step, the component calculates the common unigram between each question and answer pair and generate an AnswerScore type for the next evaluation step.
5. EvaluationDescriptor
   Token all the AnswerScore as input, calculate the prevision at N. N here is the number of true answers.

2 Third-part package

I used Stanford-corenlp (1.3.3) for tokenization.

3 Scoring Function

In ScoreDescriptor, I calculate the ratio between common unigram number and total unigram number of the answer.

4 Design Highlight

One problem in ScoreDescriptor is that for each Token or unigram, we have to know which answer or question it belongs to. In my design, I add an FSArray of tokens as a feature, so that we can easily compare with the tokens with each question and answer to calculate the score.