HW3-Report

Kartik Goyal
kartikgo@cs.cmu.edu

October 8, 2013

1 CPE execution

A CPE contains 3 main components; the collection reader, the analysis engine, and the cas consumer.
In this homework, the requirement is to handle only file system reader. Hence, I used the implementation from the UIMA example project named FileSystemCollectionReader, which converts each file in a file system into a CAS.
To create a consumer, the final engine in the homework2 had to be modified to turn it inot a CAS consumer. The steps are:

- Change the extended parent class to CasConsumer ImplBase.
- The 'process' method is changed to 'processCas' and the parameter is changed to type 'CAS'.The exception is also different.
- Get the JCas from the Cas provided in the argument and leave the code unchanged.
- Replace the destroy() method by collectionProcessComplete() method.

2 UIMA-AS

I downloaded and installed the UIMA-AS libraries correctly. I also set UIMA_PATH and UIMA_CLASSPATH correctly.
Then I ran startBroker.sh with the brokerURL mentioned in the client adn deployment descriptors.

2.1 scnlp

I added the recommended dependencies and prepared a stanford nlp client.
My original engine gave a precision of 0.58 based on token,POS,lemma and ngram overlaps.
However, the addition of NamedEntityMention annotations from scnlp hurt my performance and the precision was 0.42.
This is because of the added weight to the named entities which were common among all the answers and the questions.
I also added my own named entities by locally running stanford corenlp to my engine. Hence, now my local engine performs with precision of 0.42.
Addition of scnlp named entities keep the F-score same.
The time taken by scnlp CPE was 6007 ms with 1500ms being dedicated to the network overhead.

The processing times of my local engine and scnlp client CPE were similar around 3000 ms.
Hence, I observed that using client lowed things down becuase of the network anf the I/O overhead.
Also, better features like semantic role labels and agents and patients are better suited for this task.

2.2 UIMA AS Deployment

I created client and deployment scripts for my hw2 engine, now augmented with local named entities. The performance is exactly the same as running the engine locally on my machine. In a nutshell, my engine calculates score based on the following:

- Tokens
- Named Entities
- Part of Speech tags
- Lemma

I had to copy the maven dependencies into a folder. This is in my target directory. Hence I added this folder and my target dorectory in the UIMA classpath.

2.3 bonus

I have compared the speed and performance of the scnlp tagged and locally tagged system.
Also, POS tags and lemma are added from Stanford core nlp.