CS563 - NLP

(Read all the instructions carefully and adhere to them.)

Assignment - 2: Coreference Resolution

Deadline: March 20, 2020

Design a Coreference Resolution system to identify the masked name from the sentence.

Input: A sentence and candidate names

Example: Gina arrives and she is furious with Denise for not protecting Jody from Kingsley, as [MASK] was meant to be the parent.

Candidates: Gina, Denise

Output: Correct name

Denise

Approach:

- Solve the problem by designing a decision tree classifier (you can use the existing implementation) with features.
- You can create a positive example by replacing [MASK] with the correct name and negative example by replacing [MASK] with the other name from the candidate list.
- You have to implement a minimum of 6 meaning-full features from the following papers:
    - https://www.aclweb.org/anthology/J01-4004.pdf
    - https://www.aclweb.org/anthology/P02-1014.pdf

Dataset:

- Download the dataset from here.
- Train your system with ***WikiCREM_train.txt.*** You can use ***WikiCREM_dev.txt*** to evaluate system performance.
- In both train and dev file:
    - Each example is given in 5 lines.
    - The first line is the sentence, with one noun replaced with [MASK].

- The second line is [MASK] (the word that has to be replaced).
- The third line contains both candidates, separated with a comma. Note that the order of the candidates is NOT guaranteed to be random.
- The fourth line contains the correct candidate.
- The fifth line is empty.

Evaluation:

- Overall precision, recall and F1-score

Submission guidelines:

- Please adhere to the following guidelines while submitting your assignment.
- Please submit your assignment **on or before the deadline**.
- Compress all your files **(Input / Output / Codes / Analysis)** in zip file. It should be named as **Roll_number-Assignment-#.zip**
- Please submit your assignment on "https://www.dropbox.com/request/Z5hddljE2hxRiLfJjdkp**".**