

Authorship Detection

Minutes for 12th meeting on 07th April 2011

Held in Innova Lunch Area third floor, 4.00PM – 5.00PM

Participants: Brian, Maryam, Clement, Joel, Jie Dong

Absent: Derek

Minutes

1. Project status/progress during last week

- a. the functions of each algorithm was described again to refresh the memory.
- b. English text was verified to be all right and it was the encoding problem of the notepad.
- c. Trigram Markov
- d. Word Recurrence Interval
 - i. Results was obtained using the Federalist Text.
 - ii. 15 training data, 12 disputed text which suppose to be written by Madison.
 - iii. Results shown was not favourable. The prediction was very inconsistent, achieving a low accuracy rate of 53% most of the time.
 - iv. it was suggested that the testing could be bias to Madison as only Madison text was taken as testing data.
 - v. Comparing with the earlier results using the English text, which involves 100 training data and 70 disputed text, the accuracy and consistency was even much lower.
 - vi. WRI might not be suitable for authorship detection.
 - vii. It was suggested to combine the Function Word Frequency developed by Joel to enhance the algorithm.
- e. Function word Frequency

2. Project goals for upcoming week

- a) Clement:
 - i. do another testing.
 - ii. implement enhanced version of the WRI.
 - iii. prepare powerpoint slides for the seminar.
 - iv. start to make the video.
- b) Joel:
 - i. do another testing.
 - ii. implement enhanced version of the WRI.
 - iii. prepare powerpoint slides for the seminar.
 - iv. start to make the video.
 - v.
- c) Dong Jie
 - i. do another testing.
 - ii. implement enhanced version of the WRI.
 - iii. prepare powerpoint slides for the seminar.
 - iv. start to make the video.
 - v.

