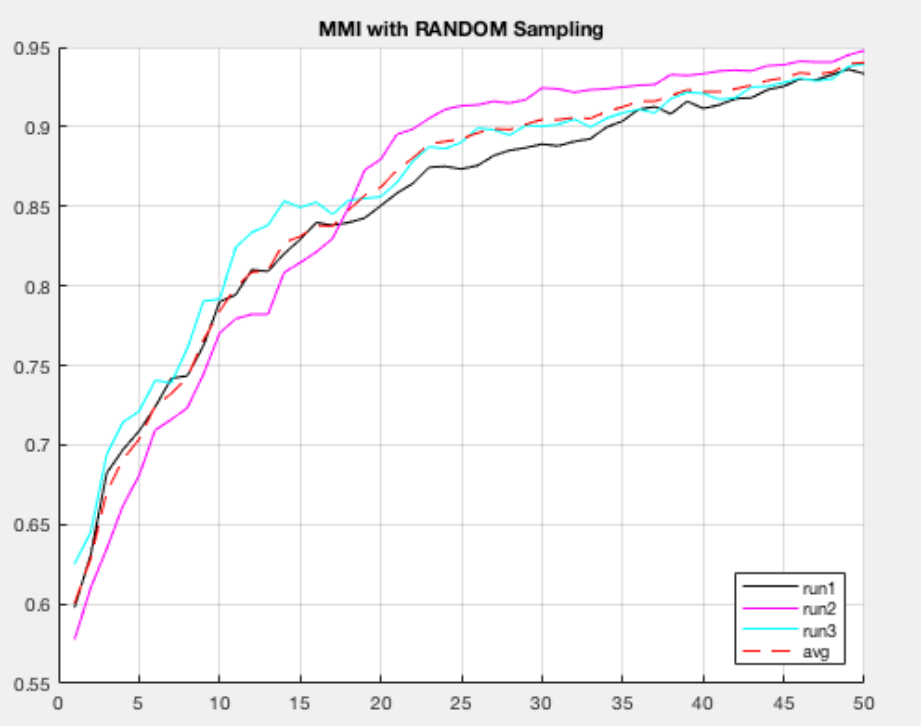
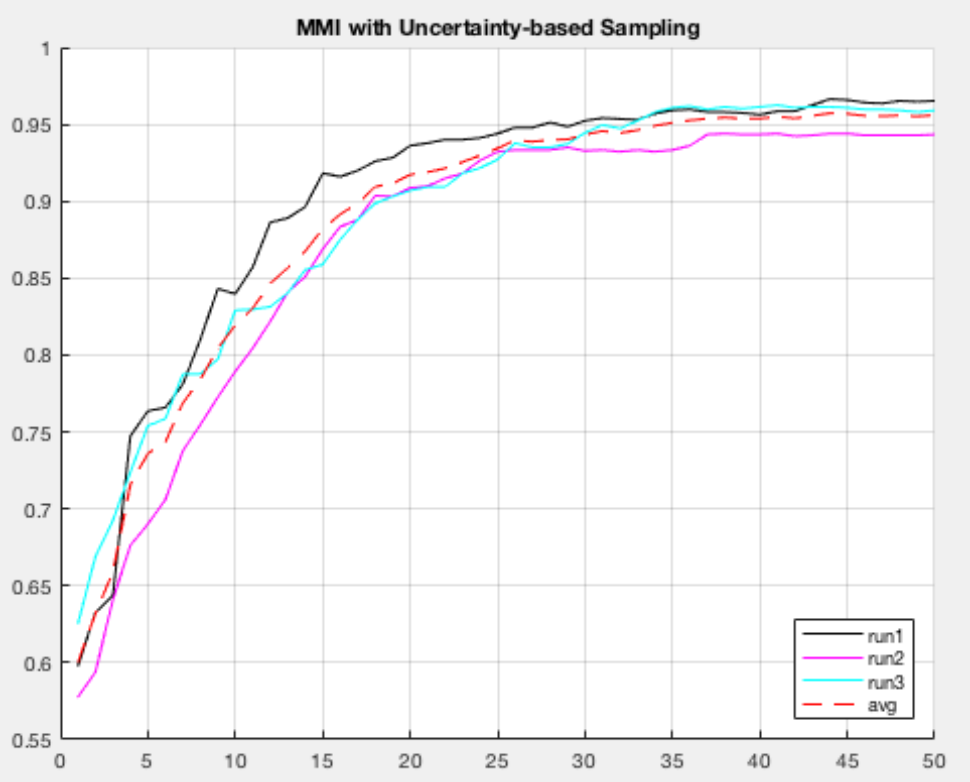
Data Mining Fall 2017

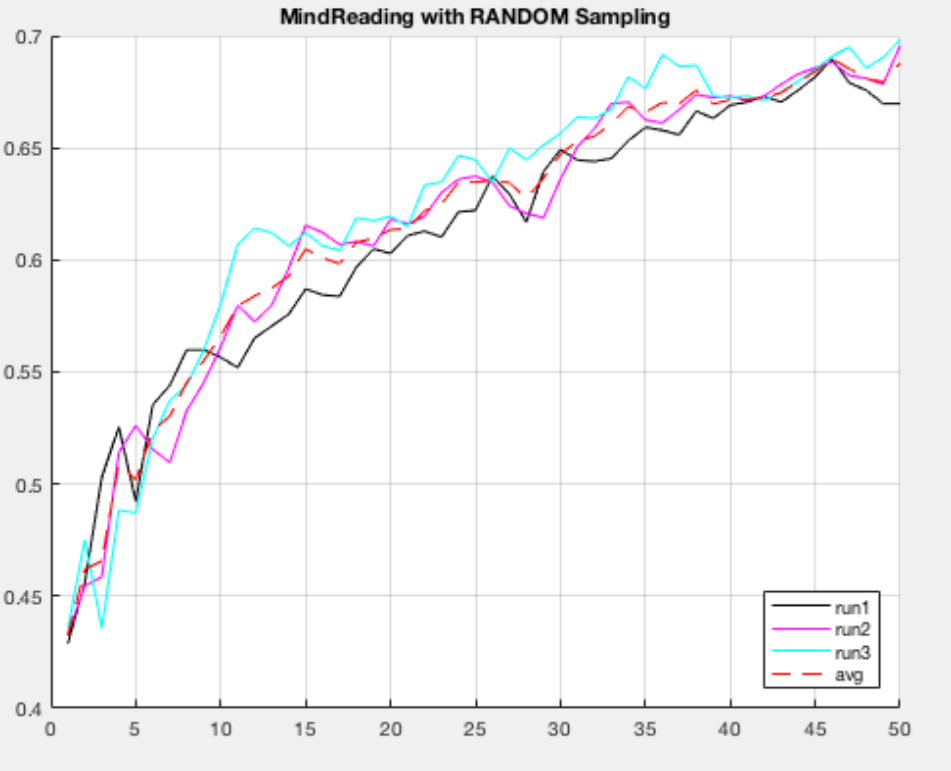
- Group Members:
  - 1209564366, Han Yu
  - 1211181735, Ying-Jheng Chen

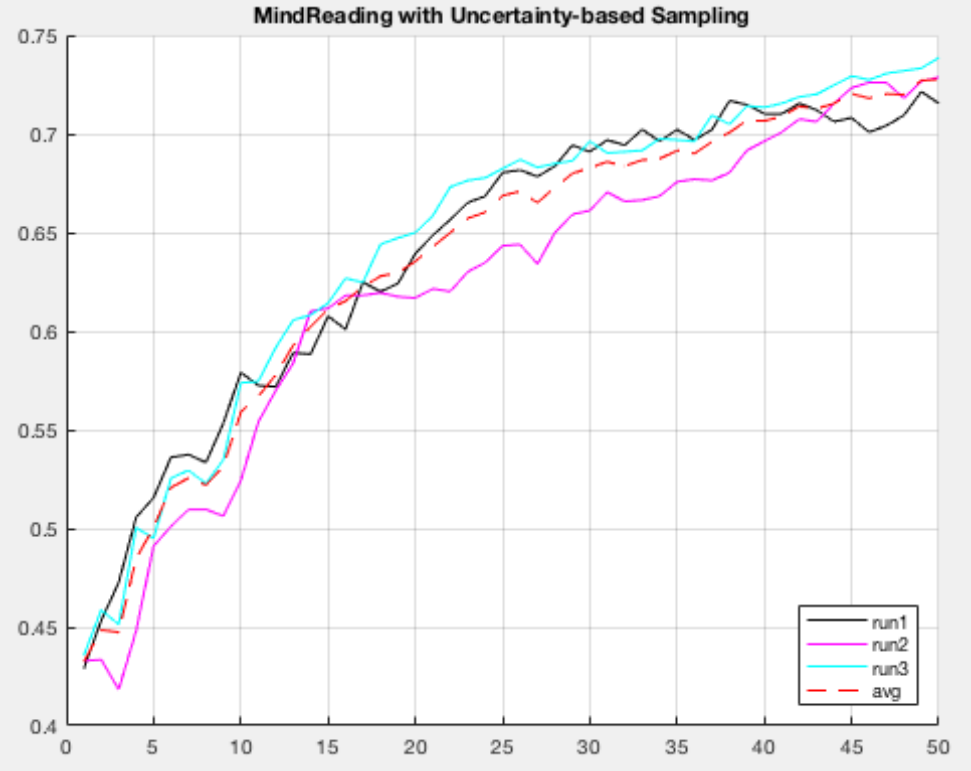
- Running Environment:
  - OS: macOS Sierra version 10.12.13
  - MacBook Pro (Retina, 13-inch, Late 2013)
  - Processor: 2.4 GHz, Intel Core i5
  - Memory: 8GB 1600 MHz DDR3
  - SSD 256 GB
  - Graphics: Intel Iris 1536 MB

- Problem 1:
  - Execution Output:
    - If k = 3, average SSE in K-menas method is 5.879040e+02
    - If k = 5, average SSE in K-menas method is 4.035847e+02
    - If k = 7, average SSE in K-menas method is 3.039352e+02
  - By SSE, we can discover that the easily possible way to improve performance is to increase the number of cluster
- Problem 2:
  - Observation:
    - With more iterations, the learning methods can have better performance.
    - Obviously, when we use Uncertainty-based Sampling, we can have better performance.
  - The following four figures are the output of data sets with different sampling methods




MMI with Uncertainty-based Sampling
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0
5
10
15
20
25
30
35
40
45
50
run1
run2
run3
avg


MindReading with RANDOM Sampling
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0
5
10
15
20
25
30
35
40
45
50
run1
run2
run3
avg


MindReading with Uncertainty-based Sampling
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0
5
10
15
20
25
30
35
40
45
50
run1
run2
run3
avg