

HW3 Report:**1. Turn Index:**

To find this I have used one Java code findTurn.java in which I counted counts for each speaker and ration to total number of counts. Higher the number means higher the involvement.

2. POS Count:

To find this I have used Java code findPOS.java in which I have calculated POS counts. Higher the number means higher the involvement.

3. Positive Negative language ratio:

For this, first I have extracted all the sentences for each user in different files using java code. Then I used python code which uses nltk library to calculate whether it is positive or negative. Then I calculated ratio positive to negative and the higher the ratio means higher the involvement.

4. Distance between language vectors:

For this, I have used java code to find number of distinct words and its count in whole dialog and for each and every user. Then I used Euclidian distance to find vector. The lower the vector means higher the involvement.

Dataset: 2009_04_28_annotated.xml

Calculated features:

Names	Truth	Turn Index	POS	Positive Negative	Vector
meg	1	1	2	1	2
amy	2	2	1	4	1
mara	3	3	4	3	3
michael	4	4	3	2	1
george	5	5	5	3	4
nick	6	6	6	6	6
michelle	7	7	7	7	7

Correlation and calculated weight:

	Correlation	Weight
Turn Index	1	0.280899
POS	0.928571429	0.258427
Positive Negative	0.79385662	0.22191
Vector	0.846227921	0.235955

Final result with ground truth:

Names	Truth	Turn Index	POS	Positive Negative	Vector	Total
meg	1	0.28	0.5	0.22	0.46	1.46
amy	2	0.56	0.25	0.88	0.23	1.92
mara	3	0.84	1	0.66	0.69	3.19
michael	4	1.12	0.75	0.44	0.23	2.54
george	5	1.4	1.25	0.66	0.92	4.23
nick	6	1.68	1.5	1.32	1.38	5.88
michelle	7	1.96	1.75	1.54	1.61	6.86