CSI 660: Social Computing
Final Project
Report

By,

Ketul Patel

(**kmpatel@albany.edu**)

Project Title: Leaderseek

Aim: To design a system that will identify the leader from the conversation available in text format like chat from an online forum, online chat room, etc.

Dataset used: I have used dataset provided by Professor for homework assignment-1 having file names 2009_04_08.txt and 2009_05_06.txt.

Implementation of the Project:

- I have converted given dataset files into xml files using standard NLP core suite.
- I have decided leadership by following measures:
  - ➢ Topic Control
  - ➢ Task Control
  - ➢ Disagreement
  - ➢ Involvement
- Detailed description of how I calculated these measures and what tools were used is as follow:

1. Topic Control
I have used following indices to calculate topic control measure:
- Turn Length
- Local Topic Introduction

Turn Length was calculated as average number of words per turn for each speaker. For finding turn length, I have used script_TurnLength.py script.

Local Topic Introduction proportion of local topics introduced by each participant, by counting the number of first mentions of local topics by each participant as percentage of all local topics in a discourse. For finding Local Topic Introduction, I have used Topic_Control.py script.

2. Task Control
I have used following indices to calculate task control measure:
- Directive Index (DI)

Directive Index was calculated as the number of directives and offers, i.e., utterances where one speaker asks or directs another, or offers/commits self to do something, made by each participant as a percentage of all directives and offers in discourse., I have used New_DA.py script.

I have used nps_chat() from nltk.corpus. I have trained model using their dataset and classified my dataset to dialogue act tagging using SK Learn classifier. All of the code can be observed in New_DA.py

3. Disagreement

I have used Disagreement.py script to calculate the measure of disagreement which is CDM (Cumulative Disagreement Measure).

4. Involvement

I have used following indices to calculate topic control measure:

- Turn Index

Turn Index was calculated as the ratio of turns per participant to the total number of turns in the discourse.

How I found leadership behaviors for both data sets:

- As I have mentioned above, leadership has been decided by these 4 measures.
- Participants have been ranked on the relative degree to which they exhibit those social behaviors.
- As mentioned in the papers Modeling Leadership and Influence in Multi-Party Online Discourse Weighting scheme for the English chat dialogue is:

Leader score = $(\alpha TCM* TCM) + (\alpha SCM* SCM) + (\alpha INVX* INVX) + (\alpha CDM* CDM)$
Where $\alpha TCM > \alpha SCM > \alpha CDM > \alpha INVX$

- Leadership With and without weights for both data sets has been shown in following tables:

➤ For 2009_04_28.txt

Name	Topic Control Measure (0.75)	Task Control Measure (0.6)	Disagreement Measure (0.3)	Involvement Measure (0.23)	Leadership with weights	Leadership Rank
Amy	0.2104	0.2272	0.2	0.2093	0.4022	1
Meg	0.1310	0.1667	0.4	0.3162	0.3909	2
George	0.1842	0.1667	0.2	0.1333	0.32882	3
Mara	0.0933	0.2575	0.2	0.1488	0.3187	4
Michelle	0.06815	0.060	0	0.03875	0.0960	7
Michael	0.2216	0.060	0	0.1147	0.2285	5
Nick	0.1160	0.060	0	0.05426	0.1355	6

➤ For 2009_05_06.txt

Name	Topic Control Measure (0.75)	Task Control Measure (0.6)	Disagreement Measure (0.3)	Involvement Measure (0.23)	Leadership with weights	Leadership Rank
Amy	0.1311	0.1538	0.125	0.1233	0.2564	4
Meg	0.2262	0.1667	0.125	0.3117	0.3788	1
George	0.2556	0.1538	0	0.1417	0.3178	3
Mara	0.1624	0.2436	0.04	0.21	0.3283	2
Michelle	0.1013	0.0256	0	0.0116	0.094003	7
John	0.069	0.1026	0	0.1133	0.1394	6
Nick	0.1082	0.1026	0.125	0.08833	0.20025	5