SOCIAL MEDIA DISCOURSE ANALYSIS

Kristine Brown, USC Information Sciences Institute
Faculty Supervisor: Jihie Kim

Background and Purpose:

There are more than 100 million Twitters users today, and this number is rapidly increasing since Twitter's inception in 2006. As more people join one of the newest forms of social media, the function of Twitter has begun to change. It is no longer just teenagers and celebrities "tweeting" (Twitter messaging) about daily life; political groups and figures have begun participating in Twitter, using it as a medium to spread their various political messages to the populations. People can reply to these tweets or "retweet" voicing their opinions and thoughts about previous tweets. Retweeting has been an important way to spread information via social media and Twitter serves as a medium for people everywhere to participate in political discussions among many other functions.

Holder (users)

Acknowledgments: Special thanks to USC's Information Sciences Institute, Dr. Jihie Kim, USC Women in Science and Engineering, and the students working on this project.

Method:

We used Linguistic Inquiry and Word Count (LIWC) to compare our human annotations. Below is LIWC vs. human annotations:

	Precision	Recall	F-score
pos	1	0.86	0.92
neg	0.97	1	1

We would try to accurately predict sentiment values using LIWC:

Target (users + entity mentioned)

	u2	u3	u5	u6	u7	u8	e1	e2	e3
u1	-	+	-	0	-	0	+	0	+
u2	0	0	0	0	0	+	0	+	0
u3	-	0	-	0	+	+	+	0	+
u4	0	+	0	+	-	+	+	+	+
u5	+	+	-	0	0	0	-	-	0
u6	-	-	-	-	-	0	0	0	0
u7	-	+	+	-	0	+	+	+	+

Note: +/- Sentiment: whether posemo is greater than negemo

Results:

Different algorithms were used to predict sentiment values. These include pure CF, PCA and PMF.
The overall accuracy of the PCA algorithm was 86% while the other target accuracy was 44%

Data Sets:

In order to find a data set with a wide range of sentiment, we focused on political discussion on Twitter split into two main groups. Discussions surrounding the terrorist group Al Shabaab in Somalia and those surrounding upcoming UK elections for prime minister. In all we collected data from more than 150,000 users and 680,000 messages.

Data Analysis:

We split up general sentiment attributes as follows:

- Positive: support, sympathy, wish-of-luck, positive-alarm
- Negative: criticize, warning, insult, frustration, ridicule
- Neutral: report, elaboration, question

Annotations were as follows

- neg/pos/neu-
- -m/g/o/p/e
  - -m means sentiment target is previous poster/message
  - The rest mean sentiment target is a group/object/person/event, respectively.

Future Work:

Add more features such as other LIWC sources and more users to build a larger foundation of data.