SENTIMENT ANALYSIS OF TWEETS
Project by
Sanmit Desai
Sindhuri Chandrupatla
INTRODUCTION

- Sentiment Analysis: refers to the use of NLP, text analysis and computational linguistics to identify and extract subjective information in source materials.

- Project Goal: Find sentiment of tweets related to Syria
DATA AND TOOLS

Data
- Pypi tool for gathering tweets
- 20k Tweets (8100 unique tweets)
- Collected tweets dated before October 2013
- Tweets related to Syria

Tools
- Java regex
- CSV reader
- JFreeChart
- Eclipse
- Python (courtesy Nilayan)
APPROACH

- For McDonald word list
  - Find a threshold and assign negative and positive values.

- Apply Affin wordlist as it is.

- Apply SentiWordNet as it is.
PROGRAMS

- **CompareSentiWordNet.java (LOC – 310)**
  - For comparing SentiWordNet with tweets

- **ReadAffinWordList.java (LOC – 92)**
  - For Affin wordList to hashmap conversion.

- **CompareAffinWordList.java (LOC – 148)**
  - For comparing Affin with tweets.

- **CompareNormalWordlist.java (LOC – 353)**
  - For comparing McDonald wordList with tweets

- **CorpusCraetor.java (LOC – 98)**
  - To separate tweets from entire corpus.

- **CovTokenizer.java (LOC – 101)**
  - To tokenize words from a sentence

- **BarChart.java (LOC – 36)**
  - To generate statistics
Starvation in Syria: a war tactic: DAMASCUS (Reuters) - One Syrian security official called it the "Starvation... http:\/\/t.co\/c0SzmuJ3Pp"
EXAMPLE TWEETS

- Polio, a once-eradicated virus, sickens 10 children and babies in war-torn Syria.
- Meet the Kurdish Female Freedom Fighters of Syria http:\/\/t.co\u002Fu0oN3dUiW8
- #Syria #FSA "WE NEVER GIVE UP! Our ideals are pure, sanctify and full of hope. God Save Syria! nhttps:\/\/t.co\lIHbAPtYIs3
**ARCHITECTURE**

- **Corpus**
- **Tweets**

**Sentiment Analyzer**
- Get Individual Tweets
- Calculate Sentiment score

**Sentiment of tweets**
**Sentiment Analyzer**

1. Separate Sentence into words
2. Compare words with the wordlist (Hashmap)
3. Add the sentiment values
4. Sentiment of a tweet
**Example Sentiment Score Calculation**

- It was a good match but an unsatisfactory result.

<table>
<thead>
<tr>
<th>Word</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT</td>
<td>0</td>
</tr>
<tr>
<td>WAS</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>GOOD</td>
<td>+1</td>
</tr>
<tr>
<td>MATCH</td>
<td>0</td>
</tr>
<tr>
<td>BUT</td>
<td>0</td>
</tr>
<tr>
<td>AN</td>
<td>0</td>
</tr>
<tr>
<td>UNSATISFACTORY</td>
<td>-2</td>
</tr>
<tr>
<td>RESULT</td>
<td>0</td>
</tr>
</tbody>
</table>

Total Sentiment = +1 -2 = -1
Populate Hashmap with +ve and –ve words

Find the frequency of each word from both lists

Sort the words by frequencies

Calculate the threshold

Values>=thre
= 1 (positive list)
= -1 (negative list)

Values < thre
= 2 (positive list)
= -2 (negative list)
Sentiment With Normal Word List

![Sentiment Graph](attachment:image.png)
AFINN

- AFINN is a list of English words rated for valence with an integer between minus five (negative) and plus five (positive).

- The words have been manually labeled by Finn Årup Nielsen in 2009-2011.

- A new evaluation of a word list for sentiment analysis in microblogs.
Sentiment With Affin Word List

![Sentiment Graph](image)
**SentiWordNet**

- SentiWordNet is sentiment analysis lexical resource made up of synset from WordNet, a thesaurus-like resource; they are allocated a sentiment score of positive, negative or objective.

- These scores are automatically generated using the semi-supervised method.

- Each term in WordNet database is assigned a score of 0 to 1 in SentiWordNet which indicates its polarity.

- Strong partiality information terms are assigned with higher scores whereas less bias/subjective terms carry low scores.
Sentiment With SentiWordNet

Number of Tweets

Sentiment Score

0 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000 5,500 6,000 6,500 7,000 7,500 8,000 8,500

Sentiment Graph
CHALLENGES

- Deciding the threshold in McDonald word list.
- Consideration of retweet.
- Twitter Rate Limit.
- Corpus processing.
- Comparison with other pre-existing wordlist.
  - Each wordlist as a different format.
  - Reading and populating wordlist in the code.

- How to generate the output (GUI)
REFERENCES

- Sentiwordnet (http://sentiwordnet.isti.cnr.it/)
- Affin (http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010)
- McDonald WordList (http://www3.nd.edu/~mcdonald/Word_Lists.html)
- Wikipedia
QUESTIONS ?

Thank You !!