Identifying Purpose in Electoral Tweets

Saif Mohammad, Svetlana Kiritchenko, Joel Martin
National Research Council Canada

Clouds of tweet words @MittRomney and @BarackObama (Oct 23 - Nov 7)
Applications of Purpose Detection in Tweets

- Determining political alignment of the electorate
- Identifying controversial issues and political opinions
- Summarizing the stream of political tweets
Collecting Election Tweets

- Polled the Twitter API for certain hashtags
  - August—September 2012

<table>
<thead>
<tr>
<th>Hashtag</th>
<th>Hashtag</th>
<th>Hashtag</th>
</tr>
</thead>
<tbody>
<tr>
<td>#4moreyears</td>
<td>#Barack</td>
<td>#campaign2012</td>
</tr>
<tr>
<td>#dems2012</td>
<td>#democrats</td>
<td>#election</td>
</tr>
<tr>
<td>#election2012</td>
<td>#gop2012</td>
<td>#gop</td>
</tr>
<tr>
<td>#joebiden2012</td>
<td>#mitt2012</td>
<td>#Obama</td>
</tr>
<tr>
<td>#ObamaBiden2012</td>
<td>#PaulRyan2012</td>
<td>#president</td>
</tr>
<tr>
<td>#president2012</td>
<td>#Romney</td>
<td>#republicans</td>
</tr>
<tr>
<td>#RomneyRyan2012</td>
<td>#veep2012</td>
<td>#VP2012</td>
</tr>
<tr>
<td>Barack</td>
<td>Obama</td>
<td>Romney</td>
</tr>
</tbody>
</table>

- Number of tweets: about one million
  - most frequent: #election2012, #campaign, #gop, #president

- Removed: non-english tweets, badly spelled tweets, retweets
Amazon’s Mechanical Turk

- **Requester**
  - breaks task into small independent units – **HITs**
  - specifies:
    - compensation for solving each HIT

- **Turkers**
  - attempt as many HITs as they wish
Two Phases of Annotation

- Questionnaire I:
  - Identifies tweets with emotion
  - Determine style and purpose of tweet
  - Determines if tweet is relevant to 2012 US elections

Annotated about 2000 tweets.
Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q1. Which of the following best describes the Emotions in this tweet? (required)

- This tweet has no emotional content.
- There is some emotion here, but the tweet does not give enough context to determine which emotion it is.
- This tweet expresses or suggests an emotional attitude or response to something.
- This tweet expresses or suggests two or more contrasting emotional attitudes or responses. (For example, the tweeter likes X but dislikes Y and Z.)
- It is not possible to decide which of the above options is appropriate because of reasons such as: the tweet does not give enough information, one needs additional context to understand the emotion, and the tweet does not make sense because of weird spellings.
Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q1. Which of the following best describes the Emotions in this tweet? (required)
- This tweet has no emotional content.
- There is some emotion here, but the tweet does not give enough context to determine which emotion it is. 87.98%
- This tweet expresses or suggests an emotional attitude or response to something.
- This tweet expresses or suggests two or more contrasting emotional attitudes or responses. (For example, the tweeter likes X but dislikes Y and Z.)
- It is not possible to decide which of the above options is appropriate because of reasons such as: the tweet does not give enough information, one needs additional context to understand the emotion, and the tweet does not make sense because of weird spellings.
Questionnaire

Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q1. Which of the following best describes the Emotions in this tweet? (required)

- This tweet has no emotional content. 8.21%
- There is some emotion here, but the tweet does not give enough context to determine which emotion it is.  
- This tweet expresses or suggests an emotional attitude or response to something. 87.98%
- This tweet expresses or suggests two or more contrasting emotional attitudes or responses. (For example, the tweeter likes X but dislikes Y and Z.)
- It is not possible to decide which of the above options is appropriate because of reasons such as: the tweet does not give enough information, one needs additional context to understand the emotion, and the tweet does not make sense because of weird spellings.

Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q1. Which of the following best describes the Emotions in this tweet? (required)
- This tweet has no emotional content. 8.21%
- There is some emotion here, but the tweet does not give enough context to determine which emotion it is. 87.98%
- This tweet expresses or suggests an emotional attitude or response to something. 2.22%
- This tweet expresses or suggests two or more contrasting emotional attitudes or responses. (For example, the tweeter likes X but dislikes Y and Z.) 2.22%
- It is not possible to decide which of the above options is appropriate because of reasons such as: the tweet does not give enough information, one needs additional context to understand the emotion, and the tweet does not make sense because of weird spellings.
Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q2. Which of the following best describes the Style of this tweet? (required)

- simple statement or question
- exaggeration or hyperbole
- sarcasm
- rhetorical question
- understatement
- weird, surreal, or off-the-wall
- humorous, but none of the above
- none of the above

Examples of the different kinds are listed in the instructions at the top of the page.
Tweeter:  Mitt Romney is arrogant as hell. He has racism written all over his face.

Q2. Which of the following best describes the Style of this tweet? (required)

- simple statement or question  76.87%
- exaggeration or hyperbole
- sarcasm
- rhetorical question
- understatement
- weird, surreal, or off-the-wall
- humorous, but none of the above
- none of the above

Examples of the different kinds are listed in the instructions at the top of the page.
Questionnaire

Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q2. Which of the following best describes the Style of this tweet? (required)

- simple statement or question 76.87%
- exaggeration or hyperbole 9.75%
- sarcasm
- rhetorical question
- understatement
- weird, surreal, or off-the-wall
- humorous, but none of the above
- none of the above

Examples of the different kinds are listed in the instructions at the top of the page.
Questionnaire

Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q2. Which of the following best describes the Style of this tweet? (required)
- simple statement or question 76.87%
- exaggeration or hyperbole 9.75%
- sarcasm 7.39%
- rhetorical question
- understatement
- weird, surreal, or off-the-wall
- humorous, but none of the above
- none of the above

Examples of the different kinds are listed in the instructions at the top of the page.
Questionnaire

Tweeter:  Mitt Romney is arrogant as hell. He has racism written all over his face.

Q2. Which of the following best describes the Style of this tweet? (required)

- simple statement or question 76.87%
- exaggeration or hyperbole 9.75%
- sarcasm 7.39%
- rhetorical question 3.19%
- understatement
- weird, surreal, or off-the-wall
- humorous, but none of the above
- none of the above

Examples of the different kinds are listed in the instructions at the top of the page.
Q3. Which of the following best describes the Purpose of this tweet? (required)

- to point out hypocrisy or inconsistency
- to point out mistake or blunder
- to disagree
- to ridicule
- to criticize, but none of the above
- to vent
- to agree
- to praise, admire, or appreciate
- to support
- to motivate or to incite action
- to be entertaining
- to provide information without any emotional content
- none of the above
Q3. Which of the following best describes the Purpose of this tweet? (required)

- to point out hypocrisy or inconsistency
- to point out mistake or blunder
- to disagree
- to ridicule
- to criticize, but none of the above
- to vent
- to oppose

- to agree
- to praise, admire, or appreciate
- to support
- favor

- to motivate or to incite action
- to be entertaining
- to provide information without any emotional content
- other

- none of the above
Q2. Which of the following best describes the Style of this tweet? (required)

- simple statement or question
- exaggeration or hyperbole
- sarcasm
- rhetorical question
- understatement
- weird, surreal, or off-the-wall
- humorous, but none of the above
- none of the above

Word that describes style if not one of the above:

Q3. Which of the following best describes the Purpose of this tweet? (required)

- to point out hypocrisy or inconsistency
- to point out mistake or blunder
- to disagree
- to ridicule
- to criticize, but none of the above
- to vent
- to agree
- to praise, admire, or appreciate
- to support
- to motivate or to incite action
- to be entertaining
- to provide information without any emotional content
- none of the above

Q4. Is this tweet about US politics and elections? (required)

- Yes, this tweet is about US politics and elections.
- No. This tweet has nothing to do with US politics or anybody involved in it.
Q3. Which of the following best describes the Purpose of this tweet? (required)

- to point out hypocrisy or inconsistency
- to point out mistake or blunder
- to disagree
- to ridicule
- to criticize, but none of the above
- to vent

---

- to agree
- to praise, admire, or appreciate
- to support

---

- to motivate or to incite action
- to be entertaining

---

- to provide information without any emotional content
- none of the above
Q3. Which of the following best describes the Purpose of this tweet? (required)

- to point out hypocrisy or inconsistency
- to point out mistake or blunder
- to disagree
- to ridicule
- to criticize, but none of the above
- to vent

- to agree
- to praise, admire, or appreciate
- to support

- to motivate or to incite action

- to be entertaining

- to provide information without any emotional content

- none of the above

oppose: 58.07%
favor: 31.76%
other: 10.17%
Questionnaire I

Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q4. Is this tweet about US politics and elections? (required)

- Yes, this tweet is about US politics and elections.
- No. This tweet has nothing to do with US politics or anybody involved in it.
Questionnaire I

Tweeter: Mitt Romney is arrogant as hell. He has racism written all over his face.

Q4. Is this tweet about US politics and elections? (required)
- [ ] Yes, this tweet is about US politics and elections. 95.56%
- [ ] No. This tweet has nothing to do with US politics or anybody involved in it.
Agreement

- **Inter-annotator agreement (IAA)**
  - Percentage of times two annotators agree with each other.
- **Average probability of choosing the Majority Class (APMS)**

<table>
<thead>
<tr>
<th>Question</th>
<th>IAA</th>
<th>APMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Number of emotions in tweet?</td>
<td>78</td>
<td>0.85</td>
</tr>
<tr>
<td>Q2. Style of tweet?</td>
<td>56</td>
<td>0.69</td>
</tr>
<tr>
<td>Q3. Purpose of tweet: 11 fine classes</td>
<td>44</td>
<td>0.52</td>
</tr>
<tr>
<td>Q3. Purpose of tweet: 3 coarse classes</td>
<td>84</td>
<td>0.86</td>
</tr>
<tr>
<td>Q4. About 2012 US elections?</td>
<td>97</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Purpose Identification Task

11-class
- to agree
- to praise
- to support
- to point out hypocrisy
- to point out mistake
- to disagree
- to ridicule
- to criticize
- to vent
- to provide information
- none of the above

3-class
- favour
- oppose
- other

Automatically Identifying Purpose

- **Data:**
  - 1072 tweets with strong majority for 11-class task
  - 1672 tweets with strong majority for 3-class task

- **Pre-processing:**
  - URL -> http://someurl
  - UserID -> @someuser
  - Tokenization and part-of-speech (POS) tagging

- **Classifier:**
  - SVM with linear kernel

- **Experiments:**
  - 10-fold cross-validation x 10 times
Features

- **n-grams**: token n-grams, skip n-grams, character n-grams
  - Romney, Mitt_Romney, Romney_is_arrogant, …

- **part-of-speech**: # of occurrences for each POS
  - # of nouns, verbs, adjectives, …

- **word clusters**: presence of tokens from each of 1000 clusters

- **all-caps**: # of tokens with all characters in upper case
  - VOTE, PLEASE, WHEREWASFBI

- **NRC Emotion Lexicon**: # of tokens/POS/all-caps/hashtags per emotion
  - proud (anticipation, joy, trust, positive), failure (disgust, fear, sadness, negative)
Features (cont.)

- **negation**: # of negated contexts
  - It's a shame #Politics can't be honest.
  - honest -> honest_NEG, trust_E -> trust_E_NEG

- **punctuation**: # of !+, ?+,(!?)+

- **emoticons**: presence of positive/negative emoticons
  - :) , :(((

- **hashtags**: # of hashtags
  - #4moreyears, #Romney

- **elongated words**: # of words with a character repeated more than 2 times
  - soooo, yayyy
Results

Overall accuracy

<table>
<thead>
<tr>
<th></th>
<th>11-class</th>
<th>3-class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority class</td>
<td>26.49</td>
<td>58.07</td>
</tr>
<tr>
<td>SVM</td>
<td>43.56</td>
<td>73.91</td>
</tr>
</tbody>
</table>
## Results per feature group

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Accuracy</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>all features</td>
<td>43.56</td>
<td>0</td>
</tr>
<tr>
<td>all – n-grams</td>
<td>39.51</td>
<td>-4.05</td>
</tr>
<tr>
<td>all – NRC Emotion lexicon</td>
<td>42.27</td>
<td>-1.29</td>
</tr>
<tr>
<td>all – POS</td>
<td>42.63</td>
<td>-0.93</td>
</tr>
<tr>
<td>all – word clusters</td>
<td>43.24</td>
<td>-0.32</td>
</tr>
<tr>
<td>all – negation</td>
<td>43.18</td>
<td>-0.38</td>
</tr>
<tr>
<td>all – all-caps, punctuation, emoticons, hashtags, elongated words</td>
<td>43.38</td>
<td>-0.18</td>
</tr>
</tbody>
</table>
Hashtag Emotion Lexicon

- 585 emotion-related hashtags
  - #love, #annoyed, #schadenfreude
- thousands of tweets with one of these hashtags
- pointwise mutual information (PMI):

\[
PMI(\text{hashtag}, \text{word}) = \log \frac{p(\text{hashtag}, \text{word})}{p(\text{hashtag}) \cdot p(\text{word})}
\]

- add 585 features: \( \text{feature(\text{hashtag}_i)} = \sum_{w_j \in \text{tweet}} PMI(\text{hashtag}_i, w_j) \)
Results with Hashtag Lexicon

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRC Emotion lexicon</td>
<td>43.56</td>
</tr>
<tr>
<td>Hashtag lexicon</td>
<td>44.35</td>
</tr>
<tr>
<td>NRC Emotion + Hashtag lexicons</td>
<td>44.58</td>
</tr>
</tbody>
</table>
Relation: purpose and emotion

Conclusions

- Compiled a large collection of electoral tweets
- Annotated them for purpose and emotion by crowdsourcing
  - Tweeters opposed much more often than supported
- Developed SVM classifier to detect purpose
- Showed that the same emotion can be associated with different types of purpose
Future Work

- Determine the reason for emotions, when stated in tweets.
- Detect a broader range of emotions than just 8 basic ones.
- Determine intensity of emotion.
- Detect exaggeration or hyperbole.
- Analyze tweets from other domains such as natural disaster responses.
- Detect personality from tweets.

Comments and Questions!