From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales

Saif Mohammad
National Research Council Canada
Road Map

- Introduction and background
- Emotion lexicon
- Analysis of emotion words in books
Emotions

When your cartoon can get you killed

(Phil, from the San Francisco Chronicle)
speaker/writer

Death threats over South Park episode
Event

listener/reader

Extremists
Participants

Trey Parker, Matt Stone
Participants
When your cartoon can get you killed.

Words

associated with joy

When your cartoon can get you killed.

associated with sadness
Our goal

• Create a large word-emotion association lexicon through input from people.
  ◦ Examples:
    • vampire is typically associated with fear
    • startle is associated with surprise
    • bliss is associated with joy
    • death is associated with sadness
    • eager is associated with anticipation

• Use the lexicon to understand the use of emotion words in text.
Which Emotions?
Plutchik, 1980: Eight Basic Emotions

- Joy
- Trust
- Fear
- Surprise
- Sadness
- Disgust
- Anger
- Anticipation

Saif Mohammad. Tracking Emotions in Books and Mail.
Using Mechanical Turk for CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON
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
Crowdsourcing

- **Benefits**
  - Inexpensive
  - Convenient and time-saving
    - Especially for large-scale annotation

- **Challenges**
  - Quality control
    - Malicious annotations
    - Inadvertent errors
Target n-grams

Must be:
  ◦ in *Roget’s Thesaurus*
  ◦ high-frequency term in the Google n-gram corpus

Followed the Mohammad and Turney (2010) approach.
Word-Choice Question

Q1. Which word is closest in meaning to shark?
   - car  - tree  - fish  - olive

- Generated automatically
  - Near-synonym taken from thesaurus
  - Distractors are randomly chosen

- Guides Turkers to desired sense

- Aides quality control
  - If Q1 is answered incorrectly:
    - Response to Q2 is discarded
Association Questions

Q2. How much is shark associated with the emotion fear? (for example, horror and scary are strongly associated with fear)
   ◦ shark is not associated with fear
   ◦ shark is weakly associated with fear
   ◦ shark is moderately associated with fear
   ◦ shark is strongly associated with fear

• Eight such questions for the eight emotions.
• Two such questions for positive or negative.
Emotion Lexicon

- Each word-sense pair is annotated by 5 Turkers
- About 10% of the assignments were discarded due to incorrect response to Q1 (gold question)
- Targets with less than 3 valid assignments removed

NRC Emotion Lexicon
  - sense-level lexicon
    - word sense pairs: 24,200
  - word-level lexicon
    - union of emotions associated with the different senses of a word
    - word types: 14,200
MOTIVATION:
EMOTION ANALYSIS OF BOOKS
Number of Books Published in a Year (source: Wikipedia)

United Kingdom (2005) 206,000 [2]
Germany (2009) 93,124 (new titles) [6]
India (2004) 82,537 (21,370 in Hindi and 18,752 in English) [8][9]
Japan (2009) 78,555 [10]
Iran (2010) 65,000 [11][12]
Taiwan (2007) 42,018 [13]
Sources of Digitized Books

- Project Gutenberg:
  - more than 34,000 books

- Google Books Corpus (GBC):
  - 5.2 million books published from 1600 to 2009
  - English portion has 361 billion words
  - 1-grams, 2-grams, 3-grams, 4-grams, 5-grams
Applications of emotion analysis of books

- **Search**
  - Example: Which Brothers Grimm tales are the darkest?

- **Social Analysis**
  - Example: How have books portrayed entities over time? (Michel et al. 2011)

- **Literary Analysis**
  - Example: Is the distribution of emotion words in fairy tales significantly different from that in novels?

- **Summarization**
  - Example: Automatically generate summaries that capture different emotional states of characters in a novel

- **Analyzing Persuasion Tactics**
  - Example: how emotion words are used for persuasion? (Mannix, 1992; Bales, 1997)
Applications of emotion analysis of books

- Search
  - Example: Which Brothers Grimm tales are the darkest?

- Social Analysis
  - Example: How have books portrayed entities over time? (Michel et al. 2011)

- Literary Analysis
  - Example: Is the distribution of emotion words in fairy tales significantly different from that in novels?

- Summarization
  - Example: Automatically generate summaries that capture different emotional states of characters in a novel

- Analyzing Persuasion Tactics
  - Example: how emotion words are used for persuasion? (Mannix, 1992; Bales, 1997)
Saif Mohammad. Tracking Emotions in Books and Mail.
Saif Mohammad. Tracking Emotions in Books and Mail.
relative salience of trust words
relative salience of sadness words

Saif Mohammad. Tracking Emotions in Books and Mail. 23
Flow of Emotions

As You Like It

Hamlet

Saif Mohammad. Tracking Emotions in Books and Mail. 24
Saif Mohammad. Tracking Emotions in Books and Mail. 25
Emotion Word Density
average number of emotion words in every X words

Brothers Grimm fairy tales ordered as per increasing negative word density. X = 10,000.
Co-occurring Emotion Words

- Examined emotion words in proximity of target entities
- Used the Google Books Corpus
  - Looked for emotion words in 5-grams that had the target
    - Ignored emotion associated with target word
  - Grouped information into 5-year bins
Percentage of fear words in close proximity to occurrences of America, China, Germany, and India in books.
Percentage of anger words in close proximity to occurrences of man and woman in books.
Comparative Analysis

FAIRY TALES VS. NOVELS

Saif Mohammad. Tracking Emotions in Books and Mail. 30
Fairy Tales

- Archetypal characters
  - peasant, king, fairy
- Clear identification of good and bad
- Appeal through emotions (Kast, 1993, Jones 2002)
  - Convey concerns, subliminal fears, wishes, and fantasies

- Do fairy tales have higher emotion word density than novels?
- Is there a difference in the distribution of emotion words?
Corpora

- The Fairy Tale Corpus (FTC) (Lobo and Martins de Matos, 2010)
  - 453 stories
  - close to 1 million words
  - penned in the 19th century by the Brothers Grimm, Beatrix Potter, and Hans C. Andersen
  - taken from Project Gutenberg

- Corpus of English Novels (CEN) (compiled by Hendrik de Smet)
  - 292 novels written between 1881 and 1922 by 25 British and American novelists
  - 26 million words
  - taken from Project Gutenberg
Histogram of texts with different anger word densities.

On the x-axis: 1 refers to density between 0 and 100, 2 refers to 100 to 200, and so on. Density is per 10,000 words.
Histogram of texts with different joy word densities.

On the x-axis: 1 refers to density between 0 and 100, 2 refers to 100 to 200, and so on. Density is per 10,000 words.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTC</td>
<td>1417</td>
<td>467</td>
</tr>
<tr>
<td>CEN</td>
<td>1164</td>
<td>196</td>
</tr>
</tbody>
</table>
Histogram of texts with sadness word densities.

On the x-axis: 1 refers to density between 0 and 100, 2 refers to 100 to 200, and so on. Density is per 10,000 words.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTC</td>
<td>814</td>
<td>443</td>
</tr>
<tr>
<td>CEN</td>
<td>785</td>
<td>159</td>
</tr>
</tbody>
</table>

Saif Mohammad. Tracking Emotions in Books and Mail. 35
Histogram of texts with surprise word densities.

On the x-axis: 1 refers to density between 0 and 100, 2 refers to 100 to 200, and so on. Density is per 10,000 words.
Summary

- Created a large word-emotion association lexicon
- Used simple measures and visualizations to quantify and track the use of emotion words in texts
- Used the Brothers Grimm fairy tales
  - showed texts can be ordered for affect-based search
- Used the Google Books Corpus
  - tracked emotion associations of entities over time
- Used the Fairy Tales and Novels Corpora
  - showed how fairy tales tend to have more extreme emotion word densities than novels