From Once Upon a Time to Happily Ever After: Tracking Emotions in Books and Mail

Saif Mohammad
National Research Council Canada
Road Map

- Motivation and background
- Emotion Lexicon
- Analysis of emotion words in mail
- Analysis of emotion words in books

Source papers:

Crowdsourcing a Word-Emotion Association Lexicon
Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

Tracking Sentiment in Mail: How Genders Differ on Emotional Axes
Saif Mohammad and Tony Yang, In Proceedings of the ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), June 2011, Portland, OR.

From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales
Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), June 2011, Portland, OR.
Road Map

- Motivation and background
- Emotion Lexicon
- Analysis of emotion words in mail
- Analysis of emotion words in books

Source papers:

Crowdsourcing a Word-Emotion Association Lexicon
Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

Tracking Sentiment in Mail: How Genders Differ on Emotional Axes
Saif Mohammad and Tony Yang, In Proceedings of the ACL 2011 Workshop on ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), June 2011, Portland, OR.

From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales
Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), June 2011, Portland, OR.
Road Map

- Motivation and background
- Emotion Lexicon
- Analysis of emotion words in mail
- Analysis of emotion words in books

Source papers:
Crowdsourcing a Word-Emotion Association Lexicon
Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

Tracking Sentiment in Mail: How Genders Differ on Emotional Axes
Saif Mohammad and Tony Yang, In Proceedings of the ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), June 2011, Portland, OR.

From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales
Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), June 2011, Portland, OR.
Road Map

- Motivation and background
- Emotion Lexicon
- Analysis of emotion words in mail
- Analysis of emotion words in books

Source papers:
Crowdsourcing a Word-Emotion Association Lexicon
Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

Tracking Sentiment in Mail: How Genders Differ on Emotional Axes
Saif Mohammad and Tony Yang, In Proceedings of the ACL 2011 Workshop on ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), June 2011, Portland, OR.

From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales
Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), June 2011, Portland, OR.
Acknowledgments

Peter Turney
National Research Council Canada

Tony (Wenda) Yang
Simon Fraser University

Saif Mohammad. Tracking Emotions in Books and Mail. 6
Emotions

When your cartoon can get you killed

Event

Death threats over South Park episode

Speaker/writer

(Phil, from the San Francisco Chronicle)

Listener/reader

Extremists

Participants

Trey Parker, Matt Stone

Participants
When your cartoon can get you killed

associated with joy

associated with sadness

Saif Mohammad. Tracking Emotions in Books and Mail.
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
Using Mechanical Turk for CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON
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
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
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 MAIL
Applications of emotion analysis of mail

- Tracking emotions in one’s own email
- Developing a depression—happiness index
  - Example: determining risk of repeat suicide attempt
    (Osgood and Walker, 1959; Pestian et al., 2008)
- Search
  - Example: locating snippets expressing anger, for customer relation management
    (Diaz and Ruz, 2002; Dube and Maute, 1996)
- Analyzing Social Behaviour
  - Example: how genders communicate in the work place
- Analyzing Social Networks
  - Example: Tracking flow of emotions in a social network
- Developing Writing Aids
  - Example: Aids for avoiding misinterpretation
    (Liu et al., 2003)
Sources of Mail for Research

- **Enron Email Corpus** (Klimt and Yang, 2004)
  - Work-place communication
  - 200,000 emails
  - October 1998 to June 2002
  - 150 of the senior management at Enron Corporation

- **ECUE Spam Datasets**
- **Phishing Datasets**
Sources of Mail for Research (continued)

- **Suicide Notes Corpus**
  - 21 notes posted by Art Kleiner
  - I2B2 2011 will release a few thousand suicide notes

- **Hate Mail Corpus**
  - 279 pieces of hate mail sent to the *Millenium Project*

- **Love Letters Corpus**
  - 348 postings at lovingyou.com
love letters

Positive

Negative

Negative words: 1352 (22.5%)
Positive words: 4647 (77.5%)

love letters

Joy words: 1644 (23.4%)
Trust words: 1474 (21.0%)
Fear words: 400 (5.7%)
Surprise words: 619 (8.8%)
Sadness words: 400 (5.7%)
Disgust words: 284 (4.0%)
Anger words: 323 (4.6%)
Anticipation words: 855 (12.2%)
Saif Mohammad. Tracking Emotions in Books and Mail. 23
Saif Mohammad. Tracking Emotions in Books and Mail.

Relative salience of joy words

Lucky, beautiful, feeling, wonderful, hope, smile, perfect, loving, thankful, blessed, baby, true, grow, kiss, glad, treasure, marriage, gift, promise, dove, happiness, joy, laugh, music, undying, cream, sweetheart, gorgeous, inspiration, intelligence.

Bar chart showing the relative salience of joy words in love letters and hate mail.
Suicide notes - hate mail

Relative salience of disgust words

Saif Mohammad. Tracking Emotions in Books and Mail. 25
WOMEN vs. MEN: Comparative Analysis of
EMOTION WORDS IN WORK-PLACE EMAIL

Saif Mohammad. Tracking Emotions in Books and Mail. 26
Genders and Language

• Huge literature at the intersection of gender and language (see bibliographies by Schiffman, 2002 and Sunderland et al., 2002)

• Men and women use language differently, even in computer-mediated communication (Boneva et al., 2001)
  ◦ Women
    • foster personal relations (Deaux and Major, 87; Eagly and Steffen, 84)
    • share concerns and support others (Boneva et al., 2001)
  ◦ Men
    • communicate for social position (Tannen, 1991)
    • prefer to talk about activities (Caldwell and Peplau, 1982; Davidson and Duberman, 1982)

What are the differences in how genders use emotion words in work-place email?
Email Corpus

- Used the Enron Email Corpus
- Removed mails with less than 50 and more than 200 words (Cheng et al., 2009)
- Identified gender of senders by name
  - 41 female, 89 male, 20 untagged
- Mails sent by gender-unknown employees removed
- 32,045 mails remaining
  - 19,920 sent by men
  - 12,125 sent by women
emails by women - emails by men

relative salience of trust words

Saif Mohammad. Tracking Emotions in Books and Mail. 29
emails to women - emails to men

<table>
<thead>
<tr>
<th>Emotions</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>2.2</td>
</tr>
<tr>
<td>Trust</td>
<td>0.2</td>
</tr>
<tr>
<td>Fear</td>
<td>0.5</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.3</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.8</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.6</td>
</tr>
<tr>
<td>Anger</td>
<td>1.5</td>
</tr>
<tr>
<td>Anticipation</td>
<td></td>
</tr>
</tbody>
</table>

relative salience of joy words

wonderful, enthusiasm, satisfied, luck, massage, abundance, excellent, exceed, volunteer, evergreen, resources, ordination, baby, gift, child, favorite, true, buddy, tree, lover, hilarious, perfect, clown, playground, award, success, clean, holiday, merry, entertainment, blessed, joy, birth, pastor, sing, spirits, gain, diamond, electric, kiss, surprise, food, presto, favorable, beautiful, garden, excited, thanksgiving, pleasant, victory.
relative salience of anticipation words
relative salience of sadness words

Saif Mohammad. Tracking Emotions in Books and Mail. 32
men to men - women to women

relative salience of fear words
### men to men - women to women

<table>
<thead>
<tr>
<th>Emotions</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>-3.8</td>
</tr>
<tr>
<td>Trust</td>
<td>1</td>
</tr>
<tr>
<td>Fear</td>
<td>2.1</td>
</tr>
<tr>
<td>Surprise</td>
<td>-1.1</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.2</td>
</tr>
<tr>
<td>Disgust</td>
<td>1.1</td>
</tr>
<tr>
<td>Anger</td>
<td>0.5</td>
</tr>
<tr>
<td>Anticipation</td>
<td>0.1</td>
</tr>
</tbody>
</table>

#### relative salience of joy words

- hope
- improvement
- create
- resources
- electric
- promise
- respect
- holiday
- glad
- hire
- proud
- sun
- presto
- enjoy
- feeling
- joy
- pleased
- accomplish
- outstanding
- merry
- success
- white
- perfect
- accomplished
- diamond
- exciting
- excited
- luck
- inviting
- exceed
- lucky
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]

**Spain** (2008) 86,300 [7]

**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. 40
Saif Mohammad. Tracking Emotions in Books and Mail.
relative salience of trust words
relative salience of sadness words
Flow of Emotions

As You Like It

Hamlet

Saif Mohammad. Tracking Emotions in Books and Mail. 44
Brothers Grimm fairy tales ordered as per increasing negative word density. $X = 10,000$. 

Saif Mohammad. Tracking Emotions in Books and Mail. 46
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.
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 Enron Corpus
  - To show use of emotion words by men and women in a work-place setting
- 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