Emotions Evoked by Common Words and Phrases:
Using Mechanical Turk to Create an Emotion Lexicon

Saif Mohammad and Peter Turney
National Research Council Canada
Painting

The Destroyer
- Frank Frazetta

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
When your cartoon can get you killed

speaker/writer

listener/reader

Death threats over South Park episode

Event

Extremists
Participants

Trey Parker, Matt Stone
Participants

(Phil, from the San Francisco Chronicle)
Our focus: words

When your cartoon can get you killed

evokes joy

epokes sadness
Motivation for emotion detection

- Devising automatic dialogue systems that respond appropriately to different emotional states of the user.
  - customer relation models
  - intelligent tutoring systems
  - emotion-aware games
- Tracking sentiment towards politicians, movies, products.
- Determining emotional intelligence.
- Assisting in writing e-mails, documents, and other text to convey desired emotion (and avoiding misinterpretation).
- Detecting how people use emotion-bearing-words to persuade and coerce others
- Deception detection

Emotions evoked by common words and phrases. Saif Mohammad and Peter Turney.
Which emotions?

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
Base emotions

- Ekman: 6 basic emotions
  - joy, sadness, fear, anger, surprise, disgust
- Plutchik: 8
  - Ekman’s 6 + anticipation + trust
  - 4 pairs of antonymous emotions
- More proposals by Parrot, Loyban, and others
Plutchik’s wheel of emotions

- Similar emotions are adjacent
- Contrasting emotions are diametrically opposite
- The radius indicates intensity
- In the white spaces are the primary dyads – emotions that are combinations of the primary emotions
Amazon’s Mechanical Turk

- Requester
  - breaks task into small independent units – HITs
  - specifies:
    - compensation for solving each HIT
    - # of independent annotations required for each HIT
      a.k.a. # of assignments/HIT
  - uploads HITs
- Turkers
  - attempt as many HITs as they wish
- Requester
  - inspects each assignment: approves or rejects

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
Amazon’s Mechanical Turk: Features

- Inexpensive
  - $1/hour is not uncommon

- Convenient
  - Web-based
  - Scripts to upload HITs and review assignments

- Takes care of certain ethics issues
  - Anonymity
  - No pressure on workers to solve HITs
Amazon’s Mechanical Turk: Challenges

- Malicious annotations
  - Random selection or garbage data entry
  - Deliberate incorrect annotation

- Inadvertent and infrequent errors
  - Turker attempts HITs for unfamiliar words too
Emotion annotation: Challenges

- Words used in different senses and in different contexts can evoke different emotions.

  High aspect ratio wings allow low speed flight.
  The fight or flight response is crucial for survival.

- How to convey the target sense to the annotator?
  - definitions are long
  - need to discourage annotation for unfamiliar words
Our solution

Directions: Attempt HIT only if you are familiar with the word. Words in different senses may have different emotion associations. Question 1 will guide you to the intended sense.

Q1. Which word is closest in meaning (most related) to flight?
   - buying
   - avoidance
   - doubt
   - boredom

- Near-synonym is taken from a thesaurus.
  - Categories in a thesaurus act as coarse senses
- Three distracters are chosen at random
Emotion annotation: Challenges

• Words used in different senses and in different contexts can evoke different emotions.
  
  High aspect ratio wings allow low speed flight.

  The fight or flight response is crucial for survival.

• How to convey the target sense to the annotator?
  ◦ definitions are long
  ◦ need to discourage annotation for unfamiliar words

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
Identifying bad assignments

- If the word choice question is answered wrongly, then the whole assignment is discarded (answers to all questions in the HIT by the Turker are discarded)

- If an annotator gets more than 1 in 3 questions wrong, then we assume they are not following instructions.
  - We reject all their assignments.
Amazon’s Mechanical Turk: Challenges

- Malicious annotations
  - Random selection or garbage data entry
  - Deliberate incorrect annotation

- Inadvertent and infrequent errors
  - Turker attempts HITs for unfamiliar words too

Will detect it 75% of the time.
Target n-grams

- Conditions:
  - Most frequent terms in the Google n-gram corpus
  - Must be in the thesaurus in just one or two categories
- Most frequent monosemous n-grams in each of the following categories:
  - noun unigrams (200)
  - noun bigrams (200)
  - verb unigrams (200)
  - verb bigrams (200)
  - adjective unigrams (200)
  - adjective bigrams (200)
  - adverb unigrams (200)
  - adverb bigrams (200)
Target n-grams (continued)

- Most frequent monosemous terms in the General Inquirer (GI) that are:
  - marked as positive (200)
  - marked as negative (200)

- Terms in WordNet Affect Lexicon (WAL) that have one or two senses and are:
  - marked as anger terms (107)
  - marked as disgust terms (25)
  - marked as fear terms (58)
  - marked as joy terms (109)
  - marked as sadness terms (86)
  - marked as surprise terms (39)

2176 terms in all.
Questions:

1. Which word is closest in meaning (most related) to flight?
   - buying
   - avoidance
   - doubt
   - boredom

2. How positive (good, praising) is flight (for example, nice and excellent are strongly positive):
   - flight is not positive
   - flight is weakly positive
   - flight is moderately positive
   - flight is strongly positive
Questions (continued):

3. How negative (bad, criticizing) is flight
   (for example, poor and pathetic are strongly negative):
   ☐ flight is not negative
   ☐ flight is weakly negative
   ☐ flight is moderately negative
   ☐ flight is strongly negative

4. How much does flight evoke/produce the emotion joy
   (for example, happy and fun may strongly evoke joy):
   ☐ flight does not evoke joy
   ☐ flight weakly evokes joy
   ☐ flight moderately evokes joy
   ☐ flight strongly evokes joy
Numbers

- 2176 (HITs) x 5 (assignments per HIT) = 10,880 assignments
- Annotators: 1012
- Turkers spent on average about 1 minute per HIT
- Hourly wage was about $2.40 (about 4 cents per HIT)
- Total cost: US $470 (cost per term: about 22 cents)
- More than 95% of the assignments had the correct answer for the word choice question.
  - The rest were discarded.
- 2081 terms had 3 or more valid assignments
  - on average 4.75 assignments per HIT
Evocative and non-evocative

- Practical NLP applications may care for only two levels of intensity
- Example: vampire-fear

| Level            | Votes |
|------------------|-------|
| No fear          | 0     |
| Weak fear        | 1     |
| Moderate fear    | 2     |
| Strong fear      | 2     |
| Non-evocative    | 1     |
| Evocative        | 4     |

Emotions evoked by common words and phrases. 
Saif Mohammad and Peter Turney.
Percent of most frequent terms evocative of different emotions

- Anger
- Anticipation
- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Trust
- Average
- Any emotion
% of WAL anger terms evocative of different emotions as per the Turkers

Emotions evoked by common words and phrases. 
Saif Mohammad and Peter Turney.
What was missed?

baffled
covetousness
exacerbate
gravel
pesky
pestering
Anger and Joy!

adjourn
credit card
find out
gloat
spontaneously
surprised
Agreement at two intensity levels:
Majority class (m) = 3, 4, 5

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
Conclusions

- Regular folks can produce high quality emotion annotations with proper guidelines and checks:
  - Annotations match those in GI and WAL
  - High degree of agreement
    - Anticipation and trust are sources of more disagreement
- A large number of commonly used terms are evocative:
  - About 61% of the terms are evocative (evoke one or the other base emotion)
Current work

When your cartoon can get you killed, it evoked associated with joy.

When you see a baby crying, it evoked associated with sadness.
% of terms where all 5 agree

Evokes  Associated

anger  anticipation  disgust  fear  joy  sadness  surprise  trust  average

0  10  20  30  40  50  60  70  80  90
Current and future work

- Determining which terms have strong color associations and if there is a correlation with emotions.
- Determine how much near synonyms vary in emotional content.
- Empirically verify if complex emotions are indeed combinations of basic emotions.
- Create a much larger lexicon (40,000 terms, say).
  - Make lexicon publicly available.
- Use lexicon in applications.
Questions.
Comparison with GI

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
Comparison with GI

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
% of WAL joy terms evocative of different emotions as per the Turkers

Emotions evoked by common words and phrases.
Saif Mohammad and Peter Turney.
Amazon’s Mechanical Turk: Challenges

- Annotator time is precious
  - Minimum reading, minimum writing
  - Maximum information throughput

- Requestor time is precious
  - Automatic review and assimilation of annotations

**One solution**: Multiple choice questions, with examples instead of explanations.
Example question

How much does vampire evoke/produce the emotion fear? (For example, horror and scary may strongly evoke fear.)

- vampire does not evoke fear
- vampire weakly evokes fear
- vampire moderately evokes fear
- vampire strongly evokes fear