Words:
Evaluative, Emotional, Colourful, Musical!

Saif Mohammad
National Research Council Canada

Includes joint work with Peter Turney, Tony Yang, Svetlana Kiritchenko, Xiaodan Zhu, Hannah Davis, Colin Cherry, Chris Collins, and Chris Kim.
Beyond literal meaning, words have other associations that often add to their meanings.

- Associations with sentiment
- Associations with emotions
- Associations with social overtones
- Associations with cultural implications
- Associations with colours
- Associations with music
Word Associations

Beyond literal meaning, words have other associations that often add to their meanings.

- Associations with sentiment
- Associations with emotions
- Associations with social overtones
- Associations with cultural implications
- Associations with colours
- Associations with music

Connotations (or all things non-Ptydepe)
Words: Evaluative, Emotional, Colourful, Musical!
Word-Sentiment Associations

- **Adjectives**
  - *reliable* and *stunning* are typically associated with **positive** sentiment
  - *rude*, *and* *broken* are typically associated with **negative** sentiment

- **Nouns and verbs**
  - *surfing* and *promotion* typically associated with **positive** sentiment
  - *death* and *crying* typically associated with **negative** sentiment

**Goal:** Create a large word-sentiment association lexicon.
Word-Emotion Associations

Words have associations with emotions:

- attack and public speaking typically associated with fear
- yummy and vacation typically associated with joy
- loss and crying typically associated with sadness
- result and wait typically associated with anticipation

Goal: Create large word-emotion association lexicon.
Sentiment Analysis

- Is a given sentence *positive, negative, or neutral*?
- Is a word within a sentence positive, negative, or neutral?
Sentiment Analysis

- Is a given sentence positive, negative, or neutral?
- Is a word within a sentence positive, negative, or neutral?

Emotion Analysis

- What emotion is being expressed in a given sentence?
  - joy, sadness, fear, anger, guilt, pride, optimism,…
Which Emotions?

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
Plutchik, 1980: Eight Basic Emotions

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

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

- **Challenges**
  - Quality control
    - Malicious annotations
    - Inadvertent errors
  - Words used in different senses are associated with different emotions.
Word-Choice Question

Q1. Which word is closest in meaning to *cry*?
   - car   - tree   - tears   - olive

- Generated automatically
- Near-synonym taken from thesaurus
- Distractors are randomly chosen
- Guides Turkers to desired sense
- Aids quality control
- If Q1 is answered incorrectly:
  - Responses to the remaining questions for the word are discarded
Association Questions

Q2. How much is cry associated with the emotion sadness? (for example, death and gloomy are strongly associated with sadness)

- cry is not associated with sadness
- cry is weakly associated with sadness
- cry is moderately associated with sadness
- cry is strongly associated with sadness

- Eight such questions for the eight basic emotions.
- Two such questions for positive or negative sentiment.
Emotion Lexicon

- Each word-sense pair is annotated by 5 Turkers
- NRC Emotion Lexicon
  - sense-level lexicon
    - word sense pairs: 24,200
  - word-level lexicon
    - union of emotions associated with different senses
    - word types: 14,200

Available at: www.saifmohammad.com

Visualizing Emotions in Text

Papers:

- **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.
love letters

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

hate mail

Negative words: 2999 (42.4%)
Positive words: 4071 (57.6%)

suicide notes

Negative words: 228 (45.1%)
Positive words: 277 (54.9%)
relative salience of trust words
relative salience of sadness words
As You Like It

Hamlet

Frankenstein

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
Words: Evaluative, Emotional, Colourful, Musical!

A method to generate music from literature.

- music that captures the change in the distribution of emotion words.
Challenges

- Not changing existing music -- generating novel pieces
- Paralysis of choice
- Has to sound good
- No one way is the right way
- Evaluation is tricky
Music-Emotion Associations

- **Major and Minor Keys**
  - major keys: happiness
  - minor keys: sadness

- **Tempo**
  - fast tempo: happiness or excitement

- **Melody**
  - a sequence of consonant notes: joy and calm
  - a sequence of dissonant notes: excitement, anger, or unpleasantness

Pieces
Three simultaneous piano melodies pertaining to the dominant emotions.

Examples (click cover to play)

TransProse: www.musicfromtext.com
Music played 185,000 times in the last 2 months.
Words: Evaluative, Emotional, Colourful, Musical!
Word-Colour Associations

Concrete concepts

- iceberg to white
- vegetation to green

Abstract concepts

- danger to red
- honesty to white

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
Colours Add to Linguistic Information

- Strengthens the message (improves semantic coherence)
- Eases cognitive load on the receiver
- Conveys the message quickly
- Evokes the desired emotional response
Crowdsourced Questionnaire

Q. Which colour is associated with *sleep*?

- black
- green
- purple… (11 colour options in random order)

NRC Word-Colour Association Lexicon

- **sense-level lexicon**: 24,200 word sense pairs

Papers:


Visualizing Word-Colour Associations

Visualization

Chris Kim and Chris Collins, UOIT

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
<table>
<thead>
<tr>
<th>botany</th>
<th>evergreen</th>
<th>garden</th>
<th>pickle</th>
<th>sprout</th>
<th>vegetable</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 out of 8</td>
<td>12 out of 12</td>
<td>12 out of 12</td>
<td>8 out of 8</td>
<td>13 out of 13</td>
<td>8 out of 8</td>
</tr>
<tr>
<td>moss</td>
<td>weed</td>
<td>herbal</td>
<td>remittance</td>
<td>mint</td>
<td>frog</td>
</tr>
<tr>
<td>6 out of 13</td>
<td>9 out of 10</td>
<td>8 out of 9</td>
<td>8 out of 9</td>
<td>13 out of 17</td>
<td>5 out of 16</td>
</tr>
<tr>
<td>habitat</td>
<td>meadow</td>
<td>monetary</td>
<td>pay</td>
<td>wealth</td>
<td>plantation</td>
</tr>
<tr>
<td>7 out of 12</td>
<td>7 out of 8</td>
<td>7 out of 8</td>
<td>7 out of 8</td>
<td>9 out of 11</td>
<td>8 out of 10</td>
</tr>
<tr>
<td>abundance</td>
<td>cost</td>
<td>dollar</td>
<td>endeavor</td>
<td>grantee</td>
<td>leaflet</td>
</tr>
<tr>
<td>8 out of 8</td>
<td>9 out of 16</td>
<td>8 out of 8</td>
<td>9 out of 9</td>
<td>6 out of 8</td>
<td>6 out of 8</td>
</tr>
<tr>
<td>propagation</td>
<td>worth</td>
<td>mow</td>
<td>leaf</td>
<td>farm</td>
<td>garnish</td>
</tr>
<tr>
<td>9 out of 12</td>
<td>9 out of 12</td>
<td>17 out of 23</td>
<td>14 out of 19</td>
<td>16 out of 22</td>
<td>16 out of 22</td>
</tr>
<tr>
<td>amount</td>
<td>greedy</td>
<td>cultivation</td>
<td>financier</td>
<td>germ</td>
<td>lavish</td>
</tr>
<tr>
<td>22 out of 31</td>
<td>7 out of 10</td>
<td>6 out of 9</td>
<td>6 out of 9</td>
<td>19 out of 15</td>
<td>6 out of 8</td>
</tr>
<tr>
<td>payment</td>
<td>pod</td>
<td>price</td>
<td>swampy</td>
<td>weeds</td>
<td>affluence</td>
</tr>
<tr>
<td>8 out of 12</td>
<td>6 out of 9</td>
<td>14 out of 21</td>
<td>6 out of 9</td>
<td>17 out of 26</td>
<td>15 out of 23</td>
</tr>
<tr>
<td>harvest</td>
<td>restitution</td>
<td>alienation</td>
<td>army</td>
<td>camouflage</td>
<td>gain</td>
</tr>
<tr>
<td>9 out of 14</td>
<td>7 out of 11</td>
<td>16 out of 16</td>
<td>25 out of 40</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
</tr>
<tr>
<td>gainful</td>
<td>graze</td>
<td>renewal</td>
<td>teller</td>
<td>vineyard</td>
<td>envious</td>
</tr>
<tr>
<td>5 out of 8</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
</tr>
<tr>
<td>buyer</td>
<td>debenture</td>
<td>inhabit</td>
<td>assets</td>
<td>appraise</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>Relevancy</td>
<td>Count out of 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blackness</td>
<td>7</td>
<td>14 out of 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>black</td>
<td>7</td>
<td>13 out of 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evil</td>
<td>7</td>
<td>13 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thug</td>
<td>6</td>
<td>13 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>darken</td>
<td>5</td>
<td>28 out of 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>charcoal</td>
<td>4</td>
<td>25 out of 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mourn</td>
<td>3</td>
<td>10 out of 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interment</td>
<td>3</td>
<td>9 out of 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>negro</td>
<td>3</td>
<td>8 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suicidal</td>
<td>3</td>
<td>8 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pepper</td>
<td>3</td>
<td>36 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>death</td>
<td>3</td>
<td>9 out of 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>somber</td>
<td>3</td>
<td>24 out of 19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bomb</td>
<td>2</td>
<td>26 out of 22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>executioner</td>
<td>2</td>
<td>8 out of 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>perish</td>
<td>2</td>
<td>16 out of 22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subversion</td>
<td>2</td>
<td>8 out of 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blindfold</td>
<td>2</td>
<td>10 out of 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curse</td>
<td>2</td>
<td>17 out of 24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sin</td>
<td>2</td>
<td>10 out of 27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>recording</td>
<td>2</td>
<td>7 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stormy</td>
<td>2</td>
<td>7 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vulture</td>
<td>2</td>
<td>2 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adviser</td>
<td>2</td>
<td>11 out of 16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>discrimination</td>
<td>2</td>
<td>15 out of 22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>marked</td>
<td>2</td>
<td>8 out of 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sinister</td>
<td>2</td>
<td>9 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>soot</td>
<td>2</td>
<td>9 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>witch</td>
<td>2</td>
<td>9 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dire</td>
<td>2</td>
<td>21 out of 32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dislike</td>
<td>2</td>
<td>11 out of 17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>illegal</td>
<td>2</td>
<td>9 out of 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>schism</td>
<td>2</td>
<td>9 out of 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fright</td>
<td>2</td>
<td>7 out of 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>terminal</td>
<td>2</td>
<td>7 out of 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>threatening</td>
<td>2</td>
<td>7 out of 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>contraband</td>
<td>2</td>
<td>8 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deadly</td>
<td>2</td>
<td>11 out of 18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disastrous</td>
<td>2</td>
<td>24 out of 23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abomination</td>
<td>2</td>
<td>8 out of 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abyss</td>
<td>2</td>
<td>16 out of 26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>demise</td>
<td>2</td>
<td>9 out of 25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>desolation</td>
<td>2</td>
<td>6 out of 19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thief</td>
<td>2</td>
<td>12 out of 29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing</td>
<td>1</td>
<td>9 out of 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prostitute</td>
<td>1</td>
<td>6 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deepen</td>
<td>1</td>
<td>6 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reprint</td>
<td>1</td>
<td>6 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suffering</td>
<td>1</td>
<td>6 out of 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cowardly</td>
<td>nugget</td>
<td>sun</td>
<td>sunny</td>
<td>saffron</td>
<td>treasure</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>6 out of 7</td>
<td>7 out of 7</td>
<td>7 out of 7</td>
<td>9 out of 10</td>
<td>8 out of 9</td>
<td>7 out of 8</td>
</tr>
<tr>
<td>lion</td>
<td>mustard</td>
<td>radiant</td>
<td>bee</td>
<td>butter</td>
<td>insecure</td>
</tr>
<tr>
<td>6 out of 8</td>
<td>6 out of 7</td>
<td>6 out of 7</td>
<td>7 out of 13</td>
<td>11 out of 13</td>
<td>6 out of 8</td>
</tr>
<tr>
<td>sandy</td>
<td>scatter</td>
<td>lightning</td>
<td>beehive</td>
<td>practically</td>
<td>radiate</td>
</tr>
<tr>
<td>7 out of 10</td>
<td>7 out of 10</td>
<td>8 out of 11</td>
<td>10 out of 14</td>
<td>5 out of 7</td>
<td>5 out of 7</td>
</tr>
<tr>
<td>enlighten</td>
<td>sunshine</td>
<td>candlelight</td>
<td>dawn</td>
<td>honey</td>
<td>urinalysis</td>
</tr>
<tr>
<td>7 out of 10</td>
<td>7 out of 10</td>
<td>6 out of 9</td>
<td>4 out of 6</td>
<td>6 out of 9</td>
<td>6 out of 9</td>
</tr>
<tr>
<td>honeycomb</td>
<td>awaken</td>
<td>cornet</td>
<td>daze</td>
<td>kerosene</td>
<td>oriental</td>
</tr>
<tr>
<td>9 out of 14</td>
<td>7 out of 11</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
</tr>
<tr>
<td>phosphor</td>
<td>pyramid</td>
<td>quail</td>
<td>sand</td>
<td>omelet</td>
<td>breakfast</td>
</tr>
<tr>
<td>5 out of 8</td>
<td>10 out of 16</td>
<td>5 out of 8</td>
<td>5 out of 8</td>
<td>6 out of 10</td>
<td>7 out of 12</td>
</tr>
<tr>
<td>medal</td>
<td>acquainted</td>
<td>candle</td>
<td>chirp</td>
<td>conversion</td>
<td>egg</td>
</tr>
<tr>
<td>16 out of 24</td>
<td>4 out of 7</td>
<td>4 out of 7</td>
<td>8 out of 14</td>
<td>4 out of 7</td>
<td>4 out of 7</td>
</tr>
<tr>
<td>incandescent</td>
<td>innuendo</td>
<td>lessen</td>
<td>bugle</td>
<td>bus</td>
<td>coin</td>
</tr>
<tr>
<td>4 out of 7</td>
<td>4 out of 7</td>
<td>4 out of 7</td>
<td>5 out of 9</td>
<td>5 out of 9</td>
<td>5 out of 9</td>
</tr>
<tr>
<td>day</td>
<td>folly</td>
<td>hive</td>
<td>insignificant</td>
<td>lighthouse</td>
<td>plated</td>
</tr>
<tr>
<td>5 out of 9</td>
<td>5 out of 9</td>
<td>7 out of 13</td>
<td>5 out of 9</td>
<td>10 out of 18</td>
<td>5 out of 9</td>
</tr>
<tr>
<td>ripple</td>
<td>beam</td>
<td>happy</td>
<td>buzz</td>
<td>caution</td>
<td>conductivity</td>
</tr>
<tr>
<td>5 out of 9</td>
<td>6 out of 11</td>
<td>7 out of 13</td>
<td>6 out of 12</td>
<td>4 out of 8</td>
<td>4 out of 8</td>
</tr>
</tbody>
</table>

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
Words: Evaluative, Emotional, Colourful, Musical!
Hashtagged Tweets

• Hashtagged words are good labels of sentiments and emotions
  
  Some jerk just stole my photo on #tumblr #grrr #anger

• Hashtags are not always good labels:
  ◦ hashtag used sarcastically
    
    The reviewers want me to re-annotate the data. #joy
  ◦ hashtagged emotion not in the rest of the message
    
    Mika used my photo on tumblr. #anger

Generating lexicon for 500 emotions

NRC Hashtag Emotion Lexicon: About 20,000 words associated with about 500 emotions

Papers:

- **Using Nuances of Emotion to Identify Personality.** Saif M. Mohammad and Svetlana Kiritchenko, In Proceedings of the ICWSM Workshop on Computational Personality Recognition, July 2013, Boston, USA.

- **Using Hashtags to Capture Fine Emotion Categories from Tweets.** Saif M. Mohammad, Svetlana Kiritchenko, Computational Intelligence, in press.
SemEval-2013, Task 2

- Is a given message positive, negative, or neutral?
  - tweet or SMS
- Is a given term within a message positive, negative, or neutral?

International competition on sentiment analysis of tweets:
- SemEval-2013 (co-located with NAACL-2013)
- 44 teams

NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu, In Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013), June 2013, Atlanta, USA.
Sentiment Lexicons

Lists of positive and negative words.

**Positive**
- spectacular
- okay

**Negative**
- lousy
- unpredictable
Sentiment Lexicons

Lists of positive and negative words, with scores indicating the degree of association

**Positive**
- spectacular 0.91
- okay 0.3

**Negative**
- lousy -0.84
- unpredictable -0.17
Creating New Sentiment Lexicons

- Compiled a list of **seed** sentiment words by looking up synonyms of excellent, good, bad, and terrible:
  - 30 positive words
  - 46 negative words
- Polled the Twitter API for tweets with seed-word hashtags
  - A set of 775,000 tweets was compiled from April to December 2012
Automatically Generated New Lexicons

- A tweet is considered:
  - positive if it has a positive hashtag
  - negative if it has a negative hashtag

- For every word $w$ in the set of 775,000 tweets, an association score is generated:
  $$\text{score}(w) = \text{PMI}(w, \text{positive}) - \text{PMI}(w, \text{negative})$$

PMI = pointwise mutual information

If $\text{score}(w) > 0$, then $w$ is positive
If $\text{score}(w) < 0$, then $w$ word is negative
NRC Hashtag Sentiment Lexicon

- $w$ can be:
  - any unigram in the tweets: 54,129 entries
  - any bigram in the tweets: 316,531 entries
  - non-contiguous pairs (any two words) from the same tweet: 308,808 entries

- Multi-word entries incorporate context:
  - unpredictable story 0.4
  - unpredictable steering -0.7

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
Setup

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

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

- **Evaluation:**
  - Macro-averaged F-pos and F-neg
## Features

<table>
<thead>
<tr>
<th>Features</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>sentiment lexicon</td>
<td>#positive: 3, scorePositive: 2.2; maxPositive: 1.3; last: 0.6, scoreNegative: 0.8, scorePositive_neg: 0.4</td>
</tr>
<tr>
<td>word n-grams</td>
<td>spectacular, like documentary</td>
</tr>
<tr>
<td>char n-grams</td>
<td>spect, docu, visua</td>
</tr>
<tr>
<td>part of speech</td>
<td>#N: 5, #V: 2, #A:1</td>
</tr>
<tr>
<td>negation</td>
<td>#Neg: 1; ngram:perfect → ngram:perfect_neg, polarity:positive → polarity:positive_neg</td>
</tr>
<tr>
<td>word clusters</td>
<td>probably, definitely, def</td>
</tr>
<tr>
<td>all-caps</td>
<td>YES, COOL</td>
</tr>
<tr>
<td>punctuation</td>
<td>#!+: 1, #?+: 0, #!?+: 0</td>
</tr>
<tr>
<td>word clusters</td>
<td>probably, definitely, probly</td>
</tr>
<tr>
<td>emoticons</td>
<td>:D, &gt;:0</td>
</tr>
<tr>
<td>elongated words</td>
<td>soooo, yaayyy</td>
</tr>
</tbody>
</table>
NRC-Canada’s Rankings in SemEval Tasks

- SemEval-2013 Task 2: Sentiment Analysis in Twitter (40+ teams)
  - Tweets
    - message-level: 1st rank
    - term level: 1st rank
  - SMS messages
    - message-level: 1st rank
    - term level: 2nd rank

- SemEval-2014 Task 9: Sentiment Analysis in Twitter (40+ teams)
  - 1st rank in 5 of 10 subtask-dataset combinations

- SemEval-2014 Task 4: Aspect Based Sentiment Analysis (30+ teams)
  - 1st rank in two of the three sentiment subtasks and 2nd in the other

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
Feature Contributions (on Tweets)

F-scores

Words: Evaluative, Emotional, Colourful, and Musical! Saif Mohammad.
Movie Reviews

- Data from rottentomatoes.com (Pang and Lee, 2005)
- Socher et al. (2013) training and test set up
- Message-level task
  - Two-way classification: positive or negative

<table>
<thead>
<tr>
<th>System</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Majority baseline</td>
<td>50.1</td>
</tr>
<tr>
<td>(b) SVM-unigrams</td>
<td>71.9</td>
</tr>
<tr>
<td>(c) Previous best result</td>
<td>85.4</td>
</tr>
<tr>
<td>(d) Our system</td>
<td>85.5</td>
</tr>
</tbody>
</table>
Mechanical Turk Annotations

- **NRC word-emotion association lexicon**
  - Entries for 8 emotion and 2 sentiments
  - Entries for 14,200 words

- **NRC word-colour association lexicon**
  - Entries for 11 colours
  - Entries for 14,200 words

Automatically Generated Lexicons (from Tweets)

- **Hashtag Emotion Lexicon**
  - 500 emotions
  - Entries for 20,000 words

- **Hashtag Sentiment Lexicon**
  - Entries for positive and negative sentiment
  - Entries for 54,000 words

Saif Mohammad
www.saifmohammad.com
saif.mohammad@nrc-cnrc.gc.ca
Future Directions

- Bridging sentiment and emotion analysis with other modalities like music
  - Further developing the text-to-music system
  - Identifying interesting applications
- Using sentiment analysis for related problems such as stance detection
- Working not just with basic emotions but with any of the hundreds of emotions
  - Understanding the relationships between different emotions
- Developing visualizations that help understand word associations