Annotation, Modelling, Analysis of Fine-Grained Emotions on a Stance and Sentiment Detection Corpus

September 8th, 2017

Hendrik Schuff, Jeremy Barnes, Julian Mohme, Sebastian Padó, Roman Klinger
Motivation

Why criticise religions? If a path is not your own. Don't be pretentious. And get down from your throne. #religion #peace #worldpeace

Which emotions are expressed?

Anger  Anticipation  Disgust  Fear  Joy  Sadness  Surprise  Trust

XXXX  XX  X  X  X  X  X

(Self-labeling typically does not discover such cases!)
Outline

1. Motivation
2. Annotation Process and Analysis
3. Baseline Models
What is Emotion Analysis?

- **Sentiment analysis**
  - positive vs. negative (neutral, mixed)

- **Subjectivity analysis**
  - subjective vs. objective

- **Emotion analysis discrete (Ekman/Plutchik)**
  - discrete emotion classes

- **Emotion analysis continuous (Posner/Russell/Peterson)**
  - valence and arousal
Emotion Models: Ekman

- Joy
- Anger
- Disgust
- Fear
- Sadness
- Surprise
Emotion Models: Plutchik’s Wheel

- Anger
- Anticipation
- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Trust
## Previous Work and State of the Art

<table>
<thead>
<tr>
<th>Name</th>
<th>Data</th>
<th>Size</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>STS-test</td>
<td>tweets</td>
<td>498</td>
<td>General</td>
</tr>
<tr>
<td>SemEval 2013</td>
<td>tweets</td>
<td>15,196</td>
<td>General</td>
</tr>
<tr>
<td>Healthcare Reform</td>
<td>tweets</td>
<td>2,516</td>
<td>Politics</td>
</tr>
<tr>
<td>Obama-McCain Debate</td>
<td>tweets</td>
<td>3,238</td>
<td>Politics</td>
</tr>
<tr>
<td>Dialogue Earth</td>
<td>tweets</td>
<td>26,110</td>
<td>Weather/Gas</td>
</tr>
<tr>
<td>STS-GOLD</td>
<td>tweets</td>
<td>2,205</td>
<td>General</td>
</tr>
<tr>
<td>SemEval 2016</td>
<td>tweets</td>
<td>4,870</td>
<td>5 topics</td>
</tr>
<tr>
<td>Sentiment Strength</td>
<td>tweets</td>
<td>4,242</td>
<td>General</td>
</tr>
<tr>
<td>ISEAR</td>
<td>descriptions</td>
<td>7,666</td>
<td>Events</td>
</tr>
<tr>
<td>Tales</td>
<td>sentences</td>
<td>1,580</td>
<td>Fairytales</td>
</tr>
<tr>
<td>Blogs</td>
<td>blogs</td>
<td>173</td>
<td>General</td>
</tr>
<tr>
<td>SemEval 2017</td>
<td>headlines</td>
<td>1,250</td>
<td>General</td>
</tr>
<tr>
<td>WASSA Emolnt 2017</td>
<td>tweets</td>
<td>7,102</td>
<td>General</td>
</tr>
<tr>
<td>Electoral Tweets</td>
<td>tweets</td>
<td>965</td>
<td>Elections</td>
</tr>
</tbody>
</table>

⇒ No manually annotated multi-label emotion corpus of Tweets available.

(References are in the paper)
Task Description and Research Question

Corpus Generation Task

- Tweets
- Publicly available
- Multiple annotators
- Multi-label
- (Additional annotation layers available)

Research Questions

- What’s the inter-annotator agreement?
- Which annotation layers interact?
- How well is it possible to computationally estimate such annotations?
Outline

1. Motivation
2. Annotation Process and Analysis
3. Baseline Models
## Annotation Process

### Corpus

- Annotation of SemEval 2016 Twitter Corpus
  - Stance and sentiment annotations exist
  - Established corpus
  - 4,870 Tweets (from which we use 4,868 Tweets)

### Annotators

- Six annotators finished their annotations
- Minimum number of annotations per Tweet is three
- 2776 Tweets annotated by four annotators
- Undergraduate students of media-informatics
- German native speakers, college-level knowledge of English
- Training of annotators based on another set of Tweets
⇒ Seldom that all annotators agree
⇒ Some classes are more difficult (Anticipation, Disgust, Fear, Sadness, Surprise) than others (Anger, Joy, Trust)
⇒ Low number of majority vote annotations
⇒ Low quality of annotation combination?
Inter-annotator Agreement

Range of pairwise agreement between all annotation pairs

- Anger: 0.5
- Anticipation: 0.4
- Disgust: 0.3
- Fear: 0.25
- Joy: 0.5
- Sadness: 0.3
- Surprise: 0.3
- Trust: 0.5
Difficult Examples (1)

That moment when Canadians realised global warming doesn't equal a tropical vacation #BCwildfire #Canadaburns #globalwarming

Anger  Anticipation  Disgust  Fear  Joy  Sadness  Surprise  Trust
> 0.33  > 0.33  > 0.33
Difficult Examples (2)

“2 pretty sisters are dancing with cancered kid”
# Cooccurrences of Labels, ℓ=0

<table>
<thead>
<tr>
<th></th>
<th>Emotions</th>
<th>Sentiment</th>
<th>Stance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anger</td>
<td>Anticipation</td>
<td>Disgust</td>
</tr>
<tr>
<td>Anger</td>
<td>2902</td>
<td>1437</td>
<td>1983</td>
</tr>
<tr>
<td>Anticipation</td>
<td>2700</td>
<td>1016</td>
<td>1029</td>
</tr>
<tr>
<td>Disgust</td>
<td>2183</td>
<td>1024</td>
<td>512</td>
</tr>
<tr>
<td>Fear</td>
<td>1840</td>
<td>466</td>
<td>1445</td>
</tr>
<tr>
<td>Joy</td>
<td>2067</td>
<td>682</td>
<td>438</td>
</tr>
<tr>
<td>Sadness</td>
<td>2644</td>
<td>664</td>
<td>613</td>
</tr>
<tr>
<td>Surprise</td>
<td>1108</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Favor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Against</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Many cooccurrences as expected (pos w/ pos, neg w/ neg)

Positive Anger Negative Joy Positive Disgust

University of Stuttgart
Schuff, Barnes, Mohme, Padó, Klinger
September 8th, 2017

14 / 22
Examples

Positive Anger

“Lets take back our country! Whos with me? No more Democrats!2016”
“Why criticise religions? If a path is not your own. Don’t be pretentious. And get down from your throne.”

Negative Joy

“Global Warming! Global Warming! Global Warming! Oh wait, it’s summer.”
“I love the smell of Hillary in the morning. It smells like Republican Victory.”

Positive Disgust

“#WeNeedFeminism because #NoMeansNo it doesn’t mean yes, it doesn’t mean try harder!”
# Models, Experimental Setting

## MaxEnt, Linear SVM
- Bag-of-Words

## LSTM, BiLSTM
- 300 dimensional embedding
- 175 dimensional LSTM layer, 0.5 dropout rate
- 50 dimensional dense layer
- 8 output neurons

## CNN
- Convolution of window size 2,3,4
- Pooling of length 2

(Twitter specific embeddings are used.)
Models for $t=0.0$

![Graph showing performance metrics for different models with $t=0.0$.](image)
Annotation Aggregation Methods: BiLSTM

<table>
<thead>
<tr>
<th>Performance</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>64</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>0.33</td>
<td>53</td>
<td>44</td>
<td>48</td>
</tr>
<tr>
<td>0.5</td>
<td>38</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>0.66</td>
<td>38</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>0.99</td>
<td>21</td>
<td>14</td>
<td>17</td>
</tr>
</tbody>
</table>
Performance vs. Frequency

![Graph showing performance vs. frequency for different categories of emotions: Anger, Disgust, Joy, Sadness, and Surprise. The x-axis represents agreement thresholds ranging from 0.0 to 0.99, and the y-axis represents counts and performance metrics.](image-url)
Conclusion and Summary

- Multi-label emotion annotation is a challenging task
- We publish all annotations
- Aggregation by disjunction leads to annotation which can better be modeled computationally
- Linear and neural models perform similarly well
Future Work

• Develop models which take into account label interactions explicitly
• Deeper linguistic analysis of annotation properties

Availability

www.ims.uni-stuttgart.de/data/ssec
Thank you!

Availability

www.ims.uni-stuttgart.de/data/ssec

Position available in my group at Uni Stuttgart!

Ph.D. student or PostDoc in project “Structured Multi-Domain Emotion Analysis from Text”