Does Optical Character Recognition and Caption Generation Improve Emotion Detection in Microblog Posts?

June 22nd, 2017

Roman Klinger
Motivation

Happy Birthday to my incredible father @realDonaldTrump. I hope that the year to come is your best yet!

My wife says to let my #anger toward my #Lupus #FuelMyFight. It works.

I have an attitude AND IT’S NOT BECAUSE I’M A BAD PERSON, BECAUSE I WASN’T RAISED RIGHT OR BECAUSE I’M JUST MEAN. IT’S BECAUSE I FEEL AWFUL ON A DAILY BASIS.

I’m tired, I’m grumpy, and all I want to do is sleep. But, of course, my body doesn’t want me to so I have to be here, be awake, and be in pain. I’M SO IRRITATED I DON’T EVEN WANT TO DEAL WITH MYSELF SOMETIMES.

SO, PLEASE EXCUSE ME ON MY HARD DAYS. IT’S NOT YOU, IT’S NOT ME, IT’S MY BODY.
Outline

1. Motivation
2. Emotion Analysis on Tweets
3. Methods & Experimental Setting
4. Results
5. Conclusion & Discussion
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
Task Description and Research Question

Task

Given a Tweet, assign an emotion from the set:
\textit{anger, fear, joy, sadness, surprise, disgust, love, shame, trust}

Research Questions

• Which feature sets contribute to the task?
  • Tweet Text
  • Text extracted from attached image
  • Automatically generated caption for attached image

• Can we operationalize this with off-the-shelf tools?
Feature Extraction & Classifier

Tweet Text
- Bag of words, keep \[^A-Za-z0-9#\]
- Username blinding, emotion hashtags are ignored

Optical Character Recognition
- Tesseract 3.04.01
- Ignore output with less than 6 bytes

Caption Generation
- NeuralTalk 2
- Pretrained COCO data set model

Classifier:
linear maximum entropy classifier with L2 regularization
Example for Caption Generation

Ivanka Trump
@IvankaTrump

Happy Birthday to my incredible father @realDonaldTrump. I hope that the year to come is your best yet!

“a man and a woman standing next to each other”
Example for Optical Character Recognition

Motivation

Emotion Analysis on Tweets

Methods & Experimental Setting

Results

Conclusion & Discussion

Men Have Lupus

@menhave lupus

My wife says to let my #anger toward my #Lupus #FuelMyFight. It works.

I have an attitude

AND IT'S NOT BECAUSE I'M A BAD PERSON, BECAUSE I WASN'T RAISED RIGHT OR BECAUSE I'M JUST MEAN.

IT'S BECAUSE I FEEL AWFUL ON A DAILY BASIS.

I'm tired, I'm grumpy, and all I want to do is sleep. But, of course, my body doesn't want me to.

SO I HAVE TO BE HERE, BE AWAKE, AND BE IN PAIN.

I'M SO IRRITATED I DON'T EVEN WANT TO DEAL WITH MYSELF SOMETIMES.

SO, PLEASE EXCUSE ME ON MY HARD DAYS. IT'S NOT YDU, IT'S NOT ME, IT'S MY BODY.

"a picture of a man in a suit"
Corpus Generation

Data collection: self labeling

Crawl corpus from Twitter with hashtags for emotions

Sample three corpora of 200,000 Tweets

- $D_{all}$ without constraints
- $D_{OCR}$ from all instances with image and with OCR output
  ⇒ Tweets with images with text
- $D_{Vis}$ from all instances with image and without OCR output
  ⇒ Tweets with images without text

(captions are generated on all images)
## Corpus Statistics

<table>
<thead>
<tr>
<th>Emotion</th>
<th>$D_{all}$</th>
<th>w/ $\phi_{OCR}$</th>
<th>w/ $\phi_{Vis}$</th>
<th>$D_{OCR}$</th>
<th>$D_{Vis}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>joy</td>
<td>91,836</td>
<td>6,927 (8%)</td>
<td>18,672 (20%)</td>
<td>92,066</td>
<td>111,604</td>
</tr>
<tr>
<td>love</td>
<td>41,470</td>
<td>3,477 (8%)</td>
<td>8,963 (22%)</td>
<td>46,290</td>
<td>50,974</td>
</tr>
<tr>
<td>sadness</td>
<td>26,521</td>
<td>1,495 (6%)</td>
<td>2,952 (11%)</td>
<td>19,707</td>
<td>14,370</td>
</tr>
<tr>
<td>fear</td>
<td>12,721</td>
<td>1,490 (12%)</td>
<td>2,299 (18%)</td>
<td>19,400</td>
<td>7,925</td>
</tr>
<tr>
<td>anger</td>
<td>11,902</td>
<td>831 (7%)</td>
<td>1,384 (12%)</td>
<td>10,379</td>
<td>5,317</td>
</tr>
<tr>
<td>surprise</td>
<td>5,492</td>
<td>195 (4%)</td>
<td>792 (14%)</td>
<td>2,790</td>
<td>5,681</td>
</tr>
<tr>
<td>trust</td>
<td>5,170</td>
<td>561 (11%)</td>
<td>886 (17%)</td>
<td>7,274</td>
<td>3,151</td>
</tr>
<tr>
<td>shame</td>
<td>4,562</td>
<td>138 (3%)</td>
<td>193 (4%)</td>
<td>1,988</td>
<td>862</td>
</tr>
<tr>
<td>disgust</td>
<td>326</td>
<td>8 (2%)</td>
<td>22 (7%)</td>
<td>106</td>
<td>116</td>
</tr>
<tr>
<td>All</td>
<td>200,000</td>
<td>15,122 (8%)</td>
<td>36,163 (18%)</td>
<td>200,000</td>
<td>200,000</td>
</tr>
</tbody>
</table>
Experiment 1:
Tweet text features only on $D_{all}$, $D_{OCR}$, $D_{Vis}$

![Graph showing F1 scores for different emotions across datasets]

Macro-F: 71, 65, 54
Experiment 2: OCR Features on $D_{OCR}$, Caption Features on $D_{Vis}$

Macro-F: 53, 12
Feature Impact of OCR and Captions on $D_{OCR}$

$$\phi^d_{text} \cup \phi^d_{OCR} \quad \phi^d_{text} \cup \phi^d_{Vis}$$

Percentage points change in $F_1$

Emotion Analysis on Tweets

Methods & Experimental Setting

Results

Conclusion & Discussion

University of Stuttgart

Roman Klinger

June 22nd, 2017
Feature Impact of OCR and Captions on $D_{\text{Vis}}$

\[ \phi^d_{\text{text}} \cup \phi^d_{\text{OCR}} \quad \phi^d_{\text{text}} \cup \phi^d_{\text{Vis}} \]

Percentage points change in $F_1$ for $D_{\text{Vis}}$

- Joy: -8
- Love: -6
- Sadness: -4
- Fear: -2
- Anger: 0
- Shame: 0
- Surprise: 0
- Trust: 0
- Disgust: 0
- Macro-Average: 0
Conclusion & Discussion

• Optical character recognition improves emotion recognition

• Caption generation does not help

• Classification with images works surprisingly well based on Tweet text only (at least on frequent emotion classes)
Future Work

- Multiple classifiers for different object types (e.g. Caffe2 model zoo)
  - Facial emotion recognition
  - Age, gender
  - Landmarks

- Include deep learning emotion classifier taking the image as input directly

- Tweet generator instead of caption generator?