Crowd Sourcing for Sentiment Analysis

ICA 2017

http://vanatteveldt.com/ica2017-sentiment
Is this positive or negative?

Cameron urges Greek euro exit plans
Is this positive or negative?

Cameron urges Greek euro exit plans

- Cameron is positive about the "plans"
- So is he positive about a Greek Euro exit?
- So is he negative about the Greek Euro?
- Is he maybe negative about Greece?
- And is he negative about the Euro?
- And what does Greece think of all this?
What is sentiment analysis?

- Sentiment: evaluative charge of subjective language
- Subjective language: reveals the inner state of author, persons in text

- Common approaches:
  - Human coding
  - Dictionary analysis
  - Machine learning

- Applications:
  - Analyzing reviews
  - Automatic stock market trading
  - Early warning systems / flagging potential negative posts
  - ...

- E.g. Pang/Lee 2008; Liu 2012; Wiebe 2004
Why is sentiment analysis difficult?

- "Terrorist Attacks: 250 Innocent Massacred By ISIS"
- "Saddam Hussein was executed by hanging"
- "This car has better mpg than my old Volvo"
- "Preacher who applauded Orlando mass killing asked to relocate"
- "Janeane Garafalo also was an interesting character ."
- "Brexit means brexit"
- "You should read the book!"
Why is sentiment analysis difficult?

- Evaluations are inherently subjective
- Evaluative language is creative and context-sensitive
  - (even more so than factual language)
- Evaluation inherently relational, but most data/tools undirected
  - Source *likes/dislikes* target
Task/domain specific sentiment analysis

- Sentiment value of words is context dependent
- Choice of words is context dependent
- Goal is often task dependent
Crowd sourcing 101

- Design (simple) task and instructions
- Upload units to e.g. crowdflower
- Random people code your units for e.g. $.02 per unit
- Can use 'gold questions' to filter coders
Why would the crowd be good at sentiment?

- Task is "simple"
- We are (mostly) looking for a "gut feeling" response
- Subjective nature favours multiple codings
  - Spread as well as point estimate

(e.g. Benoit ea 2016; Haselmayer/Jenny 2016; Socher ea 2013)
Our approach

1. Formal codebook for expert coding [hard!]
2. Create gold standard for evaluation
3. Evaluate crowd coding on small sample
4. Run crowd coding on large corpus
5. Build and evaluate dictionary and ML model
Our approach

1. Formal codebook for expert coding [hard!]
2. Create gold standard for evaluation [hard!]
3. Evaluate crowd coding on small sample
4. Run crowd coding on large corpus
5. Build and evaluate dictionary and ML model
Our approach

1. Formal codebook for expert coding [hard!]
2. Create gold standard for evaluation [hard!]
3. Evaluate crowd coding on small sample [hard!]
4. Run crowd coding on large corpus
5. Build and evaluate dictionary and ML model
Our approach

1. Formal codebook for expert coding [hard!]
2. Create gold standard for evaluation [hard!]
3. Evaluate crowd coding on small sample [hard!]
4. Run crowd coding on large corpus
5. Build and evaluate dictionary and ML model
So what was the problem?

- We want to code relational sentiment
  - target: Sentiment about someone
  - holder: Sentiment from someone
- What are the nodes?
- How to explain difference between sentiment, target, and holder to the crowd?
Aside: creating the gold standard
Attempt 1: let crowd pick target word(s)

(1) Carefully read the following sentence:

We do not want to intimidate others, nor force our ideas on them.

(2) Please indicate who or what the opinion is about by highlighting these words in the box below.

In a sentence like “David Cameron defended his announcement”, Cameron has an opinion about “his announcement”, so highlight Don't highlight the opinion itself and don't highlight the person who expresses the opinion, so do not highlight the words “c example.

We do not want to intimidate others, nor force our ideas on them.
**Attempt 1: Problems**

<table>
<thead>
<tr>
<th>n</th>
<th>holder</th>
<th>sentiment</th>
<th>target</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>271</td>
<td>41</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>159</td>
<td>34</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>14</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>57</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>34</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>26</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

- Coders could not distinguish target, holder, sentiment
Attempt 2: select target from NPs

- Identify all NPs in the sentence
- Let crowd coder select from NP:

Please read the following sentence about the longer-term future:

Then there is the question of the longer-term future for Europe.

Is someone (or the author) positive or negative about the longer-term future in this sentence?

- No, no one is no positive or negative about the longer-term future
- Yes, someone is positive about the longer-term future
- Yes, someone is negative about the longer-term future
Attempt 2: problems

- Most NPs are not potential targets
- Overall kappa .62
- Agreement mostly on what wasn’t a target
- First filter pass to throw out ’junk’ didn’t yield enough
Attempt 3: select holder from NEs

- Can use named entity recognition to ’filter’ NPs
- Agents, organizations, locations (=countries)

Then there is the question of the longer-term future for Europe.

The country is outside the euro zone but about 40 percent of its exports go to the single currency bloc.

Two of the largest foreign investors in India are British - Vodafone and BP.

Among ministers, there was fury at the way Labour had, in their view, switched position to outflank the Government on its most vul

They speak about this shift in Brussels and Paris and certainly in Athens, Rome and Madrid.

President François Hollande also said that France could be willing to supply arms at some point.

Mr Hollande’s first official foreign trip will be to Berlin for a meeting with Ms Merkel after his inauguration on May 15.

I am not a British isolationist but I do want a better deal for Britain," he said.

Regional and Western governments have expressed growing concern about the security threat from extremists and organised crime.

- Recall of gold standard holders/targets:

<table>
<thead>
<tr>
<th>type</th>
<th>holder</th>
<th>sentiment</th>
<th>target</th>
</tr>
</thead>
<tbody>
<tr>
<td>NER</td>
<td>0.67</td>
<td>0.01</td>
<td>0.51</td>
</tr>
<tr>
<td>NP</td>
<td>0.96</td>
<td>0.45</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Please read the following sentence about India:

Two of the largest foreign investors in India are British - Vodafone and BP.

Is India positive or negative about something or someone in this sentence?

- No, India is not positive or negative about something or someone in this sentence.
- Yes, India is positive about something or someone in this sentence.
- Yes, India is negative about something or someone in this sentence.
Crowd sourcing for sentiment analysis

<table>
<thead>
<tr>
<th># of coders</th>
<th>Kappa for majority vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.60</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
</tr>
<tr>
<td>6</td>
<td>0.70</td>
</tr>
<tr>
<td>8</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Sentiment is hard

Crowd Sourcing for sentiment analysis

Coverage/kappa for holder-sentiment

Kappa

Coverage

Crowd Sourcing for Sentiment Analysis

ICA 2017
Dynamic judgments

- Some sentences are harder than others
- Crowdflower can automatically add covers if confidence is low
- Can we use this to save money?
Performance with dynamic judgments

kappa
0.60 0.65 0.70 0.75 0.80 0.85

Average coders per sentence
2 4 6 8 10

0.5
0.65
0.75
0.80
0.85

Sentiment is hard
Crowd sourcing for sentiment analysis
Second step: select target from NPs

Please read the following sentence in which France expresses a positive opinion:

President François Hollande also said that France could be willing to supply arms at some point.

About who or what is France positive?
- President François Hollande
- arms
- some point
- NONE: France is not positive about any of these
Second step: select target from NPs

- Only evaluated on 26 sentences for now
- 20 were correct, but most other reasonable
  - Problems often non-NP target ("to supply arms")
- Need to test on larger scale
Conclusion

- Sentiment is (still) hard :)
- Crowd sourcing is (still) a promising technique
- Lessons:
  - Task needs to be simple multiple choice
  - Gold questions are very important
  - Work needs to be done on generating potential holders/targets
- Plans:
  - Run larger scale evaluation
  - Crowd code larger corpus
  - Experiment with dictionary / ML