Sentiment Analysis

Recent Developments in Computational Discourse Processing

July 8, 2014

Ilya Kornev
Plan

1. **Introduction** *(Handbook of Natural Language Processing, 2nd Edition: Sentiment Analysis and Subjectivity, Bing Liu, 2010)*
   - Sentiment Analysis
   - Approaches for Sentiment Analysis

2. **Sentiment Analysis and Discourse** *(Polarity Analysis of Texts using Discourse Structure, Bas Heershop et al., 2011)*
   - RST
   - Pathos Architecture and Algorithms
   - Results
Facts vs. Opinions

Textual Information:

• Facts

   Apple released a new iPhone.

   I bought an iPhone a few days ago.

• Opinions

   It was such a nice phone.

   It is much better than that LG I had before.
Problem of Sentiment Analysis

really horrible
Author: lutfigallaro
28 June 2014

this is the really the worst movie i have seen this year. let me make this short. it is extremely boring. i wanted to get out of the theater is midway. plot is confusing. action and graphics are excellent, but what's the point when there is no story. above average acting, repeatedly irritating jokes, unnecessary music. music is good but it is added at unnecessary places. the movie could have been wrapped up in 30 min but it almost stretched for 3 hours. older actor was definitely a better actor. you won't be able to sit throughout the film if you haven't watched the other parts, but even if you have watched the other parts it won't make much of a difference. don't waste your time and money for this movie.

...THE WORST MOVIE I’VE SEEN THIS YEAR...

... IT IS EXTREMELY BORING...

...PLOT IS CONFUSING...

...ACTION AND GRAPHICS ARE EXCELLENT...
Hey GUYS! I'd like to inform you all of the KARAOKE night at Old Murpies TONIGHT. I'm sure most of you have been to this Irish Pub before. What can I say about it? Hm... GREAT ambiance, GREAT people, GREAT Location, most importantly GREAT beer! Karaoke lovers Naturally can

Or more formally: $<o_j, f_{jk}, oo_{ijkl}, h_i, t_i>$
Tasks for Sentiment Analysis

- Classification:
  - Document-level
  - Sentence-level
- Feature, opinion holder extraction
- Opinion retrieval
- Opinion spam detection
Task: determine opinion orientation of a document.
Assumption: single object, single holder.
Approaches:
• Machine Learning
• Lexicon-based approaches
Classification based on ML

Approaches: supervised and unsupervised learning

Data: product reviews with ratings (for supervised ML)

(Possible) features:

- ngrams frequencies (tf-idf)
- POS tags
- opinion words and phrases (good, wonderful, junk, cost someone an arm and a leg)
- syntactic dependencies
- negation
## Pros and Cons of Machine Learning

<table>
<thead>
<tr>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Lexicons are not crucial</td>
<td>- Need for (labeled) training data</td>
</tr>
<tr>
<td>- Very good accuracy (90-96%)</td>
<td>- Domain specific</td>
</tr>
</tbody>
</table>
The picture quality of this camera is not great [+1], but the battery life is long [0].

$$\sum = \text{not great} [-1] + \text{long} [+1] = 0$$

- Identifying opinion words and phrases
- Handling negations
- But-clauses
- Aggregating opinions
Lexicon-based Approaches: Pros and Cons

<table>
<thead>
<tr>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Less domain specific</td>
<td>• Less accurate</td>
</tr>
<tr>
<td>• Employ a more thorough linguistic analysis</td>
<td>• Don’t take into account discourse features</td>
</tr>
</tbody>
</table>
Hypothesis

Polarity Analysis of Texts using Discourse Structure, Bas Heershop et al., 2011

Try: weight words according to their position in the discourse structure

Get: increased accuracy
Rhetorical Structure Theory

RST splits a text into spans, each representing a meaningful part of the text.

Illustration taken from Manfred Stede, RST Revisited: Disentangling Nuclearity
Text Spans and their Relations

Spans:

• Nucleus
• Satellite

Relations between text spans:

• Hypotactic (nucleus - satellite)
• Paratactic (all spans are nuclei)
Relations in RST and Pathos

23 relations in RST.
8 relations in Pathos.

Although it was great to see Brad Pitt fall off a cliff, this movie was terrible.
Task: retrieve a discourse structure.

Tool: Sentence-level PArsing of DiscoursE (SPADE)

F-score = 83.1% on RST Discourse Treebank
Previous Work:
SO-CAL (Taboada et al., 2008)

Disadvantages:
- Only adjectives are considered
- Satellites are not differentiated
Pathos Architecture
Workflow

- Determine POS (Open NLP tagger)
- Lemmatize
- Perform WSD
- Assign score
- Assign weight w.r.t. the position in the discourse structure
- Aggregate word scores for each sentence
- Classify the document based on weighted average of sentence scores.
Task: assign scores to opinion words.

WordNet synsets + Positive, Objective, Negative scores
SentiWordNet and Pathos

Word Sentiment Score = Pos - Neg

-1 ≤ Word Sentiment Score ≤ 1
**Task:** assign word weight w.r.t. the position in the discourse structure.

Positioners in Pathos:
- Simple Positioner
- SPADE Positioner
- SPADE Extended Positioner
Intuition:
given the structure **Introduction - Arguments - Conclusion**, most important opinions are located towards the end of the document.

Weight Assignment:
uniformly, from 0 to 1

\[ w_1(0) \]
\[ \ldots \]
\[ w_2(1) \]
Intuition:

Opinions in nuclei are more important than in satellites.

Weight Assignment:

Word weight = span weight

Version 1:

Nucleus 1
Satellite 0

Version 2:

Nucleus 1.5
Satellite 0.5

Opinion words: adjectives, adverbs, nouns, verbs.
**Intuition:**

some satellite relations may contribute differently to the overall sentiment than others

**Weight Assignment:**

word weight = span weight

In range [-2,2]
**Solution:** Genetic Algorithm

**Genetic Algorithm:**

1. Calculate sum of scores for each RST relation.
2. Generate a set of potential solutions (chromosomes).
3. Apply tournament selection of chromosomes.
4. Choose the fittest chromosome (best $F_1$).
“Fittest Chromosome”

Weights yielded by the Genetic Algorithm.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Frequency</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleus</td>
<td>600</td>
<td>0.771</td>
</tr>
<tr>
<td>Attribution</td>
<td>461</td>
<td>0.451</td>
</tr>
<tr>
<td>Background</td>
<td>362</td>
<td>0.017</td>
</tr>
<tr>
<td>Cause</td>
<td>89</td>
<td>-0.271</td>
</tr>
<tr>
<td>Condition</td>
<td>176</td>
<td>0.304</td>
</tr>
<tr>
<td>Contrast</td>
<td>243</td>
<td>-0.660</td>
</tr>
<tr>
<td>Elaboration</td>
<td>531</td>
<td>1.400</td>
</tr>
<tr>
<td>Enablement</td>
<td>266</td>
<td>0.956</td>
</tr>
<tr>
<td>Explanation</td>
<td>134</td>
<td>-0.099</td>
</tr>
</tbody>
</table>
Aggregating Scores

Sentence:
• $score(s_i) = \sum_{w_j \in s_i} score(w_j) \times weight(w_j)$

Document:
• $eval(d) = \sum_{s_i \in d} score(s_i)$
Classifying Document

\[ class(d) = \begin{cases} 
1 & \text{if } \text{eval}(d) - \text{offset} \geq 0 \\
-1 & \text{if } \text{eval}(d) - \text{offset} < 0
\end{cases} \]

Offset: controlling bias towards positive classifications.
The Whole Picture

- POS
- Lemmatize
- WSD
- Assign scores
- Assign weights
- Aggregate word scores
- Classify the document
Evaluation

Data: collection of movie reviews (500 positive and 500 negative reviews).

Training set: 60%

Test set: 40%
## Results

Table 3: Experimental results for all positioners without offset.

<table>
<thead>
<tr>
<th>Positioner</th>
<th>Offset</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Accuracy</th>
<th>Macro $F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.000</td>
<td>0.562</td>
<td>0.775</td>
<td>0.651</td>
<td>0.637</td>
<td>0.395</td>
<td>0.488</td>
<td>0.585</td>
<td>0.569</td>
</tr>
<tr>
<td>Simple</td>
<td>0.000</td>
<td>0.581</td>
<td>0.770</td>
<td>0.662</td>
<td>0.659</td>
<td>0.445</td>
<td>0.531</td>
<td>0.608</td>
<td>0.597</td>
</tr>
<tr>
<td>SPADE I (1, 0)</td>
<td>0.000</td>
<td>0.559</td>
<td>0.810</td>
<td>0.661</td>
<td>0.655</td>
<td>0.360</td>
<td>0.465</td>
<td>0.585</td>
<td>0.563</td>
</tr>
<tr>
<td>SPADE II (1.5, 0.5)</td>
<td>0.000</td>
<td>0.570</td>
<td>0.795</td>
<td>0.664</td>
<td>0.661</td>
<td>0.400</td>
<td>0.498</td>
<td>0.598</td>
<td>0.581</td>
</tr>
<tr>
<td>SPADE Extended</td>
<td>0.000</td>
<td>0.570</td>
<td>0.810</td>
<td>0.669</td>
<td>0.672</td>
<td>0.390</td>
<td>0.494</td>
<td>0.600</td>
<td>0.582</td>
</tr>
</tbody>
</table>

Table 4: Experimental results for all positioners with offset.

<table>
<thead>
<tr>
<th>Positioner</th>
<th>Offset</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Accuracy</th>
<th>Macro $F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>2.016</td>
<td>0.687</td>
<td>0.690</td>
<td>0.688</td>
<td>0.688</td>
<td>0.685</td>
<td>0.687</td>
<td>0.688</td>
<td>0.687</td>
</tr>
<tr>
<td>Simple</td>
<td>1.290</td>
<td>0.647</td>
<td>0.660</td>
<td>0.654</td>
<td>0.653</td>
<td>0.640</td>
<td>0.647</td>
<td>0.650</td>
<td>0.650</td>
</tr>
<tr>
<td>SPADE I (1, 0)</td>
<td>1.560</td>
<td>0.668</td>
<td>0.695</td>
<td>0.681</td>
<td>0.682</td>
<td>0.655</td>
<td>0.668</td>
<td>0.675</td>
<td>0.675</td>
</tr>
<tr>
<td>SPADE II (1.5, 0.5)</td>
<td>2.796</td>
<td>0.683</td>
<td>0.690</td>
<td>0.687</td>
<td>0.687</td>
<td>0.680</td>
<td>0.683</td>
<td>0.685</td>
<td>0.685</td>
</tr>
<tr>
<td>SPADE Extended</td>
<td>2.562</td>
<td>0.732</td>
<td>0.695</td>
<td>0.713</td>
<td>0.710</td>
<td>0.745</td>
<td>0.727</td>
<td>0.720</td>
<td>0.720</td>
</tr>
</tbody>
</table>
Conclusion

- Weighting words w.r.t. discourse structure improves classification accuracy ($F_1 +4.5\%$)
- Both nuclei and satellites are relevant

**Criticisms:**
- Still only one domain
- Sentence-level RST
- Technical issues
References

• Handbook of Natural Language Processing, 2nd Edition: Sentiment Analysis and Subjectivity, Bing Liu, 2010
• Polarity Analysis of Texts using Discourse Structure, Bas Heershop et al., 2011
• Manfred Stede, RST Revisited: Disentangling Nuclearity, 2008
• http://www.sfu.ca/rst/01intro/intro.html