A Semantic Network Approach to Measuring Sentiment

James Danowski
Bei Yan
Ken Riopelle
A Sentiment Network --> SENET

Paper:
• https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7292474/

SENET Code in R available at:
• github: https://github.com/BayYan/senet
Given a corpus of text, What is the sentiment around a Target?

- Person
- Company
- Concept
- Event, etc.
Given a corpus of text, What is the sentiment around a Target?

Before
Positive +, Negative -
From Seed words to Target

After
Positive +, Negative -
From Target word to Seed words

The basic unit of analysis is the word pair in a sentence.
**SENET** Approach for Measuring Sentiment

1) **Trace:** shortest paths from sentiment words to a target word, and from a target to the sentiment words

1) **Weight:** *Inverse Square Law:* invert strength and square path length so that values are larger when closer to the target

1) **Calculate:** ratios of positivity to negativity
   - sentiment intensity
   - sentiment bias
Semantic Network Shortest Paths

Diagram showing connections between words and sentiments.
SENET – Is Target-specific Sentiment Scoring

• Measures the closeness of the sentiment words to a target word of interest.

• Improves measurement of sentiment over bag-of-words approaches such as LIWC.

• Suited to strategic communication → strengthen or weaken attributes linked with targets or introduce new ones.
Example - Semantic Network Approach
Ground Truth: Annotated Airline Tweets

• Dataset includes tweets about six US Airlines: America, Jet Blue, Southwest, United, US Airways and Virgin America, scraped during February of 2015.  
  https://www.kaggle.com/crowdflower/twitter-airline-sentiment

• 4,845 tweets, comprising 1.5 MB that human annotators classified as either positive, negative, or neutral.

• Comparison of SENET to Ground Truth: we computed the positivity ratio for negative, neutral and positive annotations of overall tweet sentiment.
Positivity/Negativity Ratio for Airline Tweets
SENET Statistics Example
Ranked by S6, Positivity Index

<table>
<thead>
<tr>
<th>Airline</th>
<th>S1 Positive Seed Count</th>
<th>S2 Negative Seed Count</th>
<th>S3 Positive Path Strength</th>
<th>S4 Negative Path Strength</th>
<th>S5 Normalized Positivity</th>
<th>S6 Positivity Index</th>
<th>S7 Positivity Intensity</th>
<th>S8 Negativity Intensity</th>
<th>S9 Normalized Positivity Index</th>
<th>S10 Normalized Intensity Index</th>
<th>Sentiment Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virgin America</td>
<td>19</td>
<td>11</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>3.14</td>
<td>6.04</td>
<td>3.32</td>
<td>4.20</td>
<td>1.27</td>
<td>0.01</td>
</tr>
<tr>
<td>Jet Blue</td>
<td>73</td>
<td>43</td>
<td>0.08</td>
<td>0.03</td>
<td>0.05</td>
<td>1.89</td>
<td>4.68</td>
<td>4.21</td>
<td>3.26</td>
<td>0.77</td>
<td>0.02</td>
</tr>
<tr>
<td>Southwest Air</td>
<td>73</td>
<td>49</td>
<td>0.07</td>
<td>0.04</td>
<td>0.05</td>
<td>1.42</td>
<td>4.55</td>
<td>4.76</td>
<td>3.17</td>
<td>0.67</td>
<td>0.02</td>
</tr>
<tr>
<td>United</td>
<td>120</td>
<td>97</td>
<td>0.11</td>
<td>0.06</td>
<td>0.08</td>
<td>1.17</td>
<td>4.07</td>
<td>4.30</td>
<td>2.83</td>
<td>0.66</td>
<td>0.01</td>
</tr>
<tr>
<td>American Air</td>
<td>67</td>
<td>68</td>
<td>0.07</td>
<td>0.04</td>
<td>0.05</td>
<td>1.15</td>
<td>4.77</td>
<td>4.10</td>
<td>3.32</td>
<td>0.81</td>
<td>0.01</td>
</tr>
<tr>
<td>US Airways</td>
<td>69</td>
<td>93</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.77</td>
<td>4.44</td>
<td>4.30</td>
<td>3.09</td>
<td>0.72</td>
<td>-0.01</td>
</tr>
</tbody>
</table>
SENET example of shortest path listings
8 possible sentiment seed-target combinations

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td></td>
<td>seed</td>
<td>target</td>
<td>path_sum</td>
<td>edge_count</td>
<td>strength</td>
<td>paths</td>
</tr>
<tr>
<td>1 Pure_Positive_From_Target_to_Seed</td>
<td>absolutely</td>
<td>virginamerica</td>
<td>10</td>
<td>3</td>
<td>1.111111111</td>
<td>virginamerica follow 3; follow southwest 3; southwest absolutely 4</td>
<td></td>
</tr>
<tr>
<td>2 Pure_Positive_From_Seed_to_Target</td>
<td>absolutely</td>
<td>virginamerica</td>
<td>11</td>
<td>3</td>
<td>1.2222222222</td>
<td>absolutely service 3; service flying 5; flying virginamerica 3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Path_Type</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Pure_Positive_From_Target_to_Seed</td>
<td></td>
</tr>
<tr>
<td>2 Pure_Positive_From_Seed_to_Target</td>
<td></td>
</tr>
<tr>
<td>3 Pure_Negative_From_Target_to_Seed</td>
<td></td>
</tr>
<tr>
<td>4 Pure_Negative_From_Seed_to_Target</td>
<td></td>
</tr>
<tr>
<td>5 Mixed_Positive_From_Target_to_Seed</td>
<td></td>
</tr>
<tr>
<td>6 Mixed_Positive_From_Seed_to_Target</td>
<td></td>
</tr>
<tr>
<td>7 Mixed_Negative_From_Target_to_Seed</td>
<td></td>
</tr>
<tr>
<td>8 Mixed_Negative_From_Seed_to_Target</td>
<td></td>
</tr>
</tbody>
</table>
Insights Gained

• Pure versus mixed sentiment paths
• Ground truth versus partial truth
• Positivity bias
• Future lexicon tuning: words that contribute to more mixed sentiment paths would be candidates for removal
**SENET Contributions**

- It measures sentiment concerning targets.
- It features a network-based approach over bag-of-words.
- The sentiment network measures produce continuous ratios, expanding the scope of useful statistical procedures.
- It better informs message design for communication campaigns.
Contacts

Jim Danowski: jdanowski@gmail.com

Bei Yan: bei.yan@stevens.edu

Ken Riopelle: kenriopelle@me.com
Thank You