Sentiment Analysis of Opinions about Self-Driving Cars

Swapneel Deshpande
Nachiket Kawitkar
Why analyze self driving car reviews?

- Self driving cars: No longer a futuristic dream
- Enable aging or disabled people to drive around
- Might affect the accident rate due to human error
- Analysis to generate a summary describing the sentiments of the buyers
- Help the skeptical buyers in deciding if they would like to invest in a self driving car
Process Flow

SAS Enterprise Miner Process Flow

Fig 1. Process Flow
Text Parsing

• Generates frequency document matrix
• Clean the unstructured data

Fig 2. Text Parsing
Text Filter

- Eliminate less frequent and irrelevant terms
- Interactive filter
- Spell check
- Concept links

Fig 3. Text Filtering
Text Filter - Synonym Grouping and Spell Check

Fig 4. Synonym Grouping in Text Filter

<table>
<thead>
<tr>
<th>TERM</th>
<th>FREQ</th>
<th># DOCS</th>
<th>KEEP</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>3748</td>
<td>3371</td>
<td>✓</td>
<td>0.015</td>
</tr>
<tr>
<td>automobile</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>automobiles</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vehicles</td>
<td>65</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cars</td>
<td>1588</td>
<td>1489</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vans</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig 5. Spell Check in Text Filter

<table>
<thead>
<tr>
<th>Parent # Docs</th>
<th>Term</th>
<th># Docs</th>
<th>Parent</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0</td>
<td>changelist</td>
<td>1.0</td>
<td>changes</td>
</tr>
<tr>
<td>24.0</td>
<td>problem</td>
<td>1.0</td>
<td>problem</td>
</tr>
<tr>
<td>4.0</td>
<td>compete</td>
<td>1.0</td>
<td>compete</td>
</tr>
<tr>
<td>24.0</td>
<td>transportation</td>
<td>1.0</td>
<td>transportation</td>
</tr>
<tr>
<td>15.0</td>
<td>bwest</td>
<td>1.0</td>
<td>best</td>
</tr>
<tr>
<td>79.0</td>
<td>autonomous</td>
<td>1.0</td>
<td>autonomous</td>
</tr>
<tr>
<td>10.0</td>
<td>crazy</td>
<td>1.0</td>
<td>crazy</td>
</tr>
</tbody>
</table>
Text Filter - Concept Links

Fig 6. Concept Link for term ‘accident’

Fig 7. Concept Link for term ‘ride’
Text Cluster

- Cluster 7 shows the excitement in people
- Cluster 6 talks about technology changes
- Cluster 4 describes the happiness of work done on this innovation
Rule Based Model

- Validation Accuracy Rate: 67.53%
- Key negative words: flaw, warn, problem, trust, scare
- Key positive words: cool, excite, wish, awesome, amaze

Fig 10. Fit Statistics for Rule Based Model
Sentiment Analysis – Statistical Model

- SAS Sentiment Analysis Studio gives a quick overview of classification of the opinions into positive and negative.
- Training: 60%, Validation: 20%, Testing: 20%
- Best Model: Smoothed Relative Frequency and No Feature Ranking model
- Overall Precision = 70.44%
- The 20% of the data for testing purpose which produces the below results for positive and negative opinions respectively.

Fig 11. Graphical results of Statistical Model
Conclusion

• Rule Based Model used for control analysis to classify positive and negative sentiments
• Sentimental Analysis Studio used for quick classification
• Concept links provide deep insights regarding relationship between unusual terms
• Can be used to analyze other products

Future Scope

• To use neutral opinions in the analysis
• To build rules so as to correctly classify sarcastic comments
• Using opinions and reviews from other social media than twitter
Swapneel Deshpande
swapned@okstate.edu
(405) 714-1241
linkedin.com/in/rightmanforthejob

Nachiket Kawitkar
kawitka@okstate.edu
(405) 762-3719
linkedin.com/in/kawitka/