Automatically Determining Review Helpfulness

Hyung Yul Choi
Advisor: Kristina Striegnitz
Motivation

➢ Too many reviews
➢ Automatically find the helpful reviews
Goal

➢ Determining the “features” of reviews
➢ Learning algorithm for prediction
Research Question

➢ What are the features of reviews that are indicative of their helpfulness?
Dataset

➢ Helpfulness Ratio = \[ \frac{\text{# of votes found helpful}}{\text{# of total votes}} \]

➢ Reviews tested for Pearson’s r
  - Have at least 10 total votes and at least 5 sentences

<table>
<thead>
<tr>
<th>Total # of Reviews</th>
<th>1241778</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Reviews, ( \geq 10 ) Votes</td>
<td>167604</td>
</tr>
<tr>
<td># of Reviews, ( \geq 10 ) Votes and ( \geq 5 ) sentences</td>
<td>116680</td>
</tr>
<tr>
<td>Average Helpfulness Ratio</td>
<td>0.78</td>
</tr>
<tr>
<td>Average Length of Review</td>
<td>108 words</td>
</tr>
</tbody>
</table>
Feature: length of review (# of words)

$r = 0.26$
Feature: Flesch-Kincaid Grade Level Test

Helpfulness Ratio vs. Readability (Flesch-Kincaid Grade)

\[ r = 0.17 \]
Feature: punctuation, exclamation mark

helpfulness ratio vs. exclamation marks

\[ r = -0.21 \]
Feature: punctuation, question mark

$r = -0.32$
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment Polarity</td>
<td>● less helpful reviews use emotionally charged language</td>
<td>(-0.15)</td>
</tr>
<tr>
<td>Number of Sentences</td>
<td>● helpful reviews are longer</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Average Sentence Length</td>
<td>● sentence length has little correlation to helpfulness</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Grammatical part-of-speech Categories</td>
<td>● noun, verb, adjective use has little correlation to helpfulness</td>
<td>(\pm 0.05)</td>
</tr>
</tbody>
</table>
Results

- Prediction model
- Random baseline accuracy: 33.3%
- Decision tree: 42.9%

### Decision Tree Confusion Matrix

<table>
<thead>
<tr>
<th>Actual</th>
<th>Poor</th>
<th>Neutral</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>308</td>
<td>146</td>
<td>133</td>
</tr>
<tr>
<td>Neutral</td>
<td>217</td>
<td>201</td>
<td>169</td>
</tr>
<tr>
<td>Good</td>
<td>192</td>
<td>187</td>
<td>208</td>
</tr>
</tbody>
</table>
Current Work

➢ Subsets of features
➢ Different # of classifications
➢ Different learning algorithms
Future Work

➢ More possible features can be explored
  ○ Lexical information
  ○ Information beyond review text
Conclusion

➢ Desire to collect helpful reviews
➢ Finding useful features
➢ Using features for helpfulness prediction