Background & Goal

• Aspect-based sentiment analysis (ABSA) plays a vital role in opinion mining
• Two kinds of aspects
  * Explicit aspect: The battery of the phone lasts many hours, so it does not need to charge frequently.
  * Implicit aspect: I don’t use it anymore as I get tired of always recharging after using just for a few hours.
• Our goal: to construct a large-scale corpus annotated with implicit aspects automatically

Automatic Construction of an Annotated Corpus with Implicit Aspects

Evaluation

• Data: 10,000 reviews about mobile phones in Amazon
• Six implicit aspects are aimed
• Promising results are obtained (except for “size”)
  * 20 to 400 sentences with implicit aspects are retrieved
  * 0.58 to 0.82 accuracy for implicit aspect identification
  * There are only a few review sentences that mention the size of mobile phones in the dataset

Proposed Method

• CRF model is trained from SemEval 2014 dataset
  • By k-means
  • Sentences are represented as vectors using SCDV
  • Number of clusters = 10% of total number of sentences
  • Cluster label = aspect
  • Most frequent aspect is chosen
  • Frequency of aspect: \( F_r(a) \)
  • Occurrence of aspect in sentences: \( O_c(a) \)
  • Discard the cluster if \( \frac{F_r(a)}{\text{# of sent. in cluster}} < T_r \)

<table>
<thead>
<tr>
<th>Implicit Aspect</th>
<th># of clusters</th>
<th># of sentences</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>battery</td>
<td>28</td>
<td>274 Explicit, 393 Implicit</td>
<td>0.82</td>
</tr>
<tr>
<td>case</td>
<td>15</td>
<td>66 Explicit, 74 Implicit</td>
<td>0.74</td>
</tr>
<tr>
<td>look(design)</td>
<td>24</td>
<td>234 Explicit, 252 Implicit</td>
<td>0.58</td>
</tr>
<tr>
<td>size</td>
<td>2</td>
<td>13 Explicit, 7 Implicit</td>
<td>0.14</td>
</tr>
<tr>
<td>screen</td>
<td>7</td>
<td>115 Explicit, 21 Implicit</td>
<td>0.76</td>
</tr>
<tr>
<td>price</td>
<td>20</td>
<td>342 Explicit, 100 Implicit</td>
<td>0.78</td>
</tr>
</tbody>
</table>