Aspect-based sentiment analysis

Method

1) Syntactic patterns of subjective phrases - matching N-grams with syntactic patterns (based on Poes [4]) and valence (based on the Daeyman lexicon [3]); no use of labels.

<table>
<thead>
<tr>
<th>N-Noun</th>
<th>N-Verb</th>
<th>V-Verb</th>
<th>Prep-Prep</th>
<th>ADJ-ADJ</th>
<th>ADJ-Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>small</td>
<td>small</td>
<td>com</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>tiny</td>
<td>tiny</td>
<td>tiny</td>
<td>com</td>
<td>very</td>
<td>very</td>
</tr>
<tr>
<td>tiny</td>
<td>tiny</td>
<td>tiny</td>
<td>com</td>
<td>very</td>
<td>very</td>
</tr>
</tbody>
</table>

2) Shallow neural classification - training a neural classifier (using fasttext with 30 epochs, learning rate of 0.1, context window of 5 and a Soft-max output function) to predict the given pros and/or cons as 'hidden word' based on the review text represented as 320-dimensional word embeddings [8].

3) Shallow neural classification with clustering - same as (2) with a generalization of the target pros and cons to reduce the long tail, by K-means clustering (K = 10).

Analysis

Baseline - Matching known pros and cons from the training set to N-grams in the test review texts.

1) Reviewer gold standard - Matching predicted pros and cons to the pros and cons put forward by the reviewer, using FuzzyWuzzy [https://github.com/seatgeek/fuzzywuzzy] scaled it (0 to similarity all at 100 (total similarity). Variants of 'none' are scored as 100 if gold standard pros / cons are empty.

2) Human assessment - Review text presented to human evaluator along with the pros and cons for one of the four systems or the reviewers pros and cons: assessment by relevance of each given pro and/or con to the review text and by completeness of these pros and cons (scaled 1 to 7). 20 reviews, 50 annotations (20%) = 100 review-system combinations, 10 annotations per combination, 20 annotations per person.

Results

Figure: Venn diagram of the overlap between predicted pros and cons by each system with the gold standard pros and cons at a similarity score of > 10.

Success depends on

• Opportunity and utility

Results

Device | pros | con | both | either | review
---|---|---|---|---|---
Smartphone | 2,282 | 666 | 3,097 | 4,575 | 10,578
Deep fryer | 1,536 | 1,012 | 2,548 | 4,575 | 7,078
Vacuum cleaner | 715 | 703 | 1,418 | 2,977 | 4,477
Espresso machine | 986 | 1,041 | 2,027 | 4,575 | 6,663
Total | 14,078 | 4,538 | 3,154 | 4,575 | 31,334

Fig. 4. Example of a review placed on kieskeurig.nl.

Table: Overview of the review data set

<table>
<thead>
<tr>
<th>Device</th>
<th>pros</th>
<th>con</th>
<th>both</th>
<th>either</th>
<th>review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone</td>
<td>2,282</td>
<td>666</td>
<td>3,097</td>
<td>4,575</td>
<td>10,578</td>
</tr>
<tr>
<td>Deep fryer</td>
<td>1,536</td>
<td>1,012</td>
<td>2,548</td>
<td>4,575</td>
<td>7,078</td>
</tr>
<tr>
<td>Vacuum cleaner</td>
<td>715</td>
<td>703</td>
<td>1,418</td>
<td>2,977</td>
<td>4,477</td>
</tr>
<tr>
<td>Espresso machine</td>
<td>986</td>
<td>1,041</td>
<td>2,027</td>
<td>4,575</td>
<td>6,663</td>
</tr>
<tr>
<td>Total</td>
<td>14,078</td>
<td>4,538</td>
<td>3,154</td>
<td>4,575</td>
<td>31,334</td>
</tr>
</tbody>
</table>

Opportunity and utility

• Only 30% of product reviews on kieskeurig.nl assigned a pro and/or con by the writer.
• They can be leveraged as distant labels to link the review text to pros and cons.
• Success depends on quality, quantity and homogeneity of the labels.
• Utility: a system that automatically extracts pros and cons from product reviews, as an input.
• Facilitate the writers of new reviews with suggested pros and cons.
• Summarize multiple reviews on the same project by the number of specific pros and cons that are mentioned.
• Provide a personalized recommendation based on valued aspects by a user.

References


Acknowledgements

This work is part of the research programmes Discussion Thread Summarization for Mobile Devices and The Automated Newsroom, which are financed by the Netherlands Organisation for Scientific Research (NWO).