Predicting the Usefulness of Amazon Reviews Using Off-The-Shelf Argumentation Mining

Marco Passon*, Marco Lippi°, Giuseppe Serra*, Carlo Tasso*

* University of Udine
° University of Modena and Reggio Emilia
Looking for a Smartphone

Amazon.com: Essential Phone in Halo Gray - 128 GB Unlocked Titanium and Ceramic phone with Edge-to-Edge Display
by Essential

1,743 customer reviews | 12 answered questions

List Price: $499.99
Price: $449.00
You Save: $50.99 (10%)

Style: Only Phone
Color: Halo Gray

- With a matte finish to both the titanium and ceramic, this Amazon exclusive Essential Phone has a more industrial look while maintaining its elegant design
- Stunning edge-to-edge Quad HD display—the largest screen-to-body ratio of any smartphone. So you get a massive screen, but on a phone that still fits comfortably in one hand
- 13MP Dual RGB + Mono Rear camera with Portrait Mode and 4K video recording. Get more natural looking shots and incredible high-resolution video
Our Assumption

● What we hope to read in a review is something that goes beyond plain option or sentiment, being rather a collection or reasons and evidence that support the overall judgment..... In short, we look for argumentative reviews

● In this work, we propose a first experimental study that aims to show how features coming from an off-the-shelf argumentation mining system can help in prediction whether a given review is useful.

● A recent work (Liu et al. 2017*) explores this assumption, but their study considers a set of 110 hotel reviews with a manual annotation of arguments

● Differently, in our work we investigate the use of features coming from an automatic system on a large publicly dataset: 117,000 Amazon Reviews.

* Haijing Liu, Yang Gao, Pin Lv, Mengxue Li, Shiqiang Geng, Minglan Li, Hao Wang, "Using Argument-based Features to Predict and Analyse Review Helpfulness", EMNLP 2017
The Proposed Approach

Product Review

Jane Morgan's unpretentious, simple style of singing appealed to me since I was a kid. She put out a lot of records, but is virtually forgotten. It's a shame, because her recordings can serve as the standard for so many modern classics. The only thing I missed on this CD was her recording of "Around the World". Other than that - - elegant perfection.

The only thing I missed on this CD was her recording of "Around the World". Other than that - - elegant perfection.

Prediction
Useful/Not Useful
MARGOT System

- MARGOT is a Websystem that performs argument mining by exploit a combination of advanced machine learning and natural language processing tecnique

- Argument Definition (same as Douglas Walton - 2009):
  - **Claim**: a concise statement that directly support or contests a topic
  - **Evidence**: segment text that supports the claim, by bringing a contribution in favour of the thesis that is contained within the claim itself.

- The system was trained on a IBM Research dataset: Debater
  - 547 Wikipedia Articles; 2294 claims and 4690 evidence fact

http://margot.disi.unibo.it/index.html
### MARGOT System

**Query document**

School violence is widely held to have become a serious problem in recent decades in many countries, especially where weapons such as guns or knives are involved. It includes violence between school students as well as physical attacks by students on school staff.

### MARGOT Pipeline:

- Each document is split into sentences.
- Each sentence is processed to produce the Constituency parse tree.
- Two classifiers, based on Tree Kernels, detect if a sentence contains claims or evidence facts.

<table>
<thead>
<tr>
<th>Claim</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score&lt;sub&gt;Claim&lt;/sub&gt;</td>
<td>Score&lt;sub&gt;Evidence&lt;/sub&gt;</td>
</tr>
<tr>
<td>Score&lt;sub&gt;Claim&lt;/sub&gt;</td>
<td>Score&lt;sub&gt;Evidence&lt;/sub&gt;</td>
</tr>
</tbody>
</table>
Our Argumentation Features

Product Review

Jane Morgan's unpretentious, simple style of singing appealed to me since I was a kid. She put out a lot of records, but is virtually forgotten. It's a shame, because her recordings can serve as the standard for so many modern classics. The only thing I missed on this CD was her recording of "Around the World". Other than that -- elegant perfection.

Argumentation features

For each category (Claim, Evidence, Argument) we compute:
- Average (3 features)
- Maximum (3 features)
- N. sentences with score > 0 (3 features)
- Percentage of sentences with score >0 (3 features)
Experimental Evaluation
Amazon Product Dataset

- **Amazon Product Dataset** contains 142.8 million of product reviews spanning May 1996 – July 2014*
- We select **three categories** (CDs and Vinyl, Electronics, TV and Movies) and we extract, for each category, **39000 reviews** having at least 75 “helpful” scores.
- A review is labeled “useful”, if the ratio between the two numbers is > 0.7

*Julian McAuley - [http://jmcauley.ucsd.edu/data/amazon/](http://jmcauley.ucsd.edu/data/amazon/)
Argumentation vs helpfulness

- Category “CDs and Vinyl” (a random subset of 200 reviews)

  - A low number of sentences that contain a claim or an evidence does not necessarily mean that the review is useless
  - A review with a high number of sentences containing a claim or an evidence is most likely a useful review
Experimental Results

The experiment has been conducted classifying reviews using:

- **M**: only argumentative features
- **BoW**: only Bag of Words features
- **Bow + M**: combination of Bag of Words and Argumentative features
- **TF-IDF**: only TF-IDF features
- **TF-IDF + M**: combination of TF-IDF and Argumentative features

**Metrics**: Accuracy (A), Precision (P), Recall (R) and F1 Score ($F_1$)

- Bag of Words/TF-IDF with argumentative features achieve the best F1 score for each category
Apple products seemed to be revered as near sacred by Gen Xers. I frankly agree that the beautiful and high-quality surfaces on Apple products is worthy of preservation. This case snaps on easily, fits perfectly, weighs little and does a great job of protecting my Macbook from scratches and mars, even on an airline security conveyor belt.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>TF-IDF</th>
<th>TF-IDF + M</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not useful</td>
<td></td>
<td></td>
<td>Useful</td>
</tr>
</tbody>
</table>

Some Examples #1
Some Examples #1

- Product Review:

Apple products seemed to be revered as near sacred by Gen Xers. I frankly agree that the beautiful and high-quality surfaces on Apple products is worthy of preservation. This case snaps on easily, fits perfectly, weighs little and does a great job of protecting my Macbook from scratches and mars, even on an airline security conveyor belt.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>TF-IDF</th>
<th>TF-IDF + M</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not useful</td>
<td>Useful</td>
<td>Useful</td>
<td></td>
</tr>
</tbody>
</table>
Some Examples #2

- Product Review:

  [...] The overrated Neil Gaiman's fantasy nightmares don't even try to make sense; pointless punches are pulled on shallow cartoon characters. The immature Doctor can't shine, stuck with griping harpies. Boo-hoo, Pond leaks. Who cares? Pond's loathsome, “Are we there yet?” of Season Five set the tone for Season Six. [...]
Some Examples #2

- Product Review:

[...] The overrated Neil Gaiman's fantasy nightmares don't even try to make sense; pointless punches are pulled on shallow cartoon characters. The immature Doctor can't shine, stuck with griping harpies. Boo-hoo, Pond leaks. Who cares? Pond's loathsome, “Are we there yet?” of Season Five set the tone for Season Six. [...]

<table>
<thead>
<tr>
<th>Prediction</th>
<th>TF-IDF</th>
<th>TF-IDF + M</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Useful</td>
<td>Not useful</td>
<td>Not useful</td>
<td></td>
</tr>
</tbody>
</table>

Note: TF-IDF technique has lower performance on long reviewers; this effect is limited by when using argumentation features. Since in this case there are not argumentation sentences, the prediction of our approach is “Not Useful”.
Some Examples #3

- Product Review:

I love this product! The price is amazing. It takes a little bit long to boot and the touch screen is a little awkward but overall AMAZING. BUY IT!!

<table>
<thead>
<tr>
<th>Prediction</th>
<th>TF-IDF</th>
<th>TF-IDF + M</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Useful</td>
<td>Not useful</td>
<td></td>
</tr>
</tbody>
</table>
Some Examples #3

- Product Review:

  I love this product! The price is amazing. It takes a little bit long to boot and the touch screen is a little awkward but overall AMAZING. BUY IT!!

<table>
<thead>
<tr>
<th>Prediction</th>
<th>TF-IDF</th>
<th>TF-IDF + M</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Useful</td>
<td>Useful</td>
<td>Not useful</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Even if there is an argumentation sentence the rest is useless.
Thanks