Aspect-based summarization of pros and cons in unstructured product reviews

Florian Kunneman[1], Sander Wubben[2], Emiel Krahmer[2] and Antal van den Bosch[1,3]

[1]Radboud University, [2]Tilburg University, [3]KNAW Meertens Institute

Aspect-based sentiment analysis

- Automatically extracting information units from a text on aspects of the target topic and their valence.
- Classically approached by means of rule-based systems or machine learning.
- We compare the two approaches applied on Dutch product reviews posted on kieskeurig.nl, leverage reviewers pros and cons as labels and evaluate both the systems and labels (original study: [2]).

Materials

Figure: Example of a review placed on kieskeurig.nl.

Device	# reviews	# pro	# con	# both	# either
Smartphone	5,133	1,007	703	697	1,012
Deep fryer	4,040	1,526	1,039	1,019	1,536
Vacuum cleaner	2,623	1,026	730	715	1,041
Espresso machine	2,282	979	682	666	986
Total	14,078	4,538	3,154	3,097	4,575

Table: Overview of the review data set

Device	# train	# dev	# test	# total
Smartphone	809	101	102	1,012
Deep fryer	1,228	154	154	1,536
Vacuum cleaner	832	104	105	1,041
Espresso machine	793	99	100	986

Table: Overview of the data used in the experimental set-up

Opportunity and utility

- Only **30%** of product reviews on kieskeurig.nl is 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, so as to:
  - **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

Method

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

Syntactic pattern	Example (Dutch)	Example English	Valence word	Pro/con
ADJ	fantastisch	fantastic	fantastic	Pro
ADJ-N	klein zuigmondje	small nozzle	small	Con
ADJ-N-N	heerlijk kopje espresso	delicious cup of espresso	delicious	Pro
ADJ-ADJ-N	onhandige extra handeling	inconvenient additional action	inconvenient	Con
ADJ-V	mooi uitgevoerd	nicely rendered	nicely	Pro
ADJ-Prep-N	goed van smaak	good flavor	good	Pro

Table: Examples of syntactic patterns extracted in the SynPat system (ADJ=Adjective, N=Noun, V=Verb, Prep=Preposition)

2) Shallow neural classification - training a neural classifier (using fast-text 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 [3].

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 ($K = 100$).

Results

Figure: Example output and evaluation

System	Predicted pros and cons	Aligned pros and cons (string matching)	Aligned pros and cons ('empty' heuristic)	# reviews aligned (% of total)	Avg. similarity ('empty')	Avg. similarity (without 'empty')
Baseline	3,877	1,469	79	432 (94%)	52.50	50.05
SynPat	1,712	1,101	83	442 (96%)	51.41	48.42
Neural	1,844	468	159	375 (81%)	58.58	44.99
Neural_clust	1,844	1,228	109	456 (99%)	48.96	44.43

Table: Results on reviewer gold standard

	Completeness ($\kappa = 0.30$)	Relevance ($\kappa = 0.56$)
Baseline	**3.90** (0.90)	0.44 (0.17)
SynPat	**4.06** (1.15)	**0.67** (0.25)
Neural	2.74 (0.86)	0.25 (0.14)
Neural_clust	2.37 (0.62)	0.18 (0.10)
Reviewers summary	**4.60** (1.13)	**0.61** (0.25)

Table: Results on human assessment

Evaluation

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 0 (no similarity at all) to 100 (total similarity). Variants of 'none' are scored as 100 if gold standard pros / cons are empty.

The role of 'None'

Many reviews include only pros or only cons. Systems that predict 'None' for pros or cons are rewarded if the reviewer omitted this evaluation as well.

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 annotators, (20*5) = 100 review-system combinations, **10 annotations per combination**, 20 annotations per person
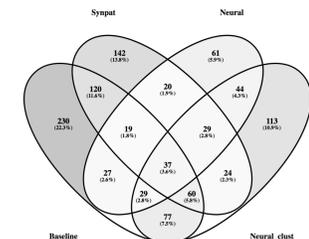
Analysis

Figure: Venn diagram of the overlap between predicted pros and cons by each system which match the gold standard pros and cons at a similarity score of > 50.

Conclusion

Reviewers' pros and cons tend to be additive and diverse; the low frequency does not help for machine learning. A generative approach to neural networks might be more suitable. SynPat shows a good performance, but is only extractive.

References

[1] Valentin Jijkoun and Katja Hofmann. Generating a non-english subjectivity lexicon: Relations that matter. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 398–405, 2009.

[2] Florian Kunneman, Sander Wubben, Antal van den Bosch, and Emiel Krahmer. Aspect-based summarization of pros and cons in unstructured product reviews. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2219–2229, 2018.

[3] Stephan Tulkens, Chris Emmery, and Walter Daelemans. Evaluating unsupervised dutch word embeddings as a linguistic resource. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, may 2016.

[4] Antal Van den Bosch, Bertjan Busser, Sander Canisius, and Walter Daelemans. An efficient memory-based morpho-syntactic tagger and parser for Dutch. In *Computational Linguistics in the Netherlands: Selected Papers from the Seventeenth CLIN Meeting*, pages 99–114, 2007.

Contact Information

- Web: http://http://discosumo.ruhosting.nl/
- Email: f.kunneman@let.ru.nl