NLP Project: Feature Extraction from Customer Reviews

Esra Gücükbel

June 15, 2021

1 Goal & Topic

When people buy a product, mostly have tendencies to look at the reviews of buyers who have been already bought same product to decide. The aim of this project is to extract Pros and Cons of the products from customer review data and show mostly showed up properties of them.

2 Dataset

In this study, Amazon Arts, Crafts and Sewing category dataset and its metadata will be used. [1]

The review dataset includes **overall,vote,verified,reviewTime,reviewerId,asin, style, reviewerName,reviewText,summary,unixReviewTime, image** attributes, however only **overall,reviewText,summary, asin** will meet the requirements for the purpose of this project.

From the metadata, only **title,main_cat,brand,asin, price** attributes will be used. Metadata and review data set is joined on **asin** attribute. (*Note:* In accordance with memory limits, main_cat,brand and price columns can be removed.)

	overall	asin	style	reviewText	summary
0	5.0	0449819906	{'Format:': ' Kindle Edition'}	I've read this book already and I've got plans...	A WONDERFUL BOOK
1	5.0	0449819906	{'Format:': ' Kindle Edition'}	Nicely written directions.	Nice
2	5.0	0449819906	{'Format:': ' Paperback'}	love it	Five Stars
3	5.0	0449819906	{'Format:': ' Kindle Edition'}	Good additional knitting reference to have ava...	Good Reference in Kindle Edition
4	5.0	0449819906	{'Format:': ' Kindle Edition'}	A gazillion pattern stitches, lucidly explaine...	Extremely clear, thorough

Figure 1: First 5 records of review dataset.

	title	main_cat	brand	asin	price
0	You Son of a Bitch! 1987 Embroidered Patch	<img src="https://images-na.ssl-images-amazon....	Honchosfx	6665560953	$6.41
1	Origami Stars Papers Package 1H (5 packs)	Books		7000000376	
2	Yi De Ge Chinese Calligraphy Sumi Drawing Blac...	Arts, Crafts & Sewing	MasterChinese	7000001089	
3	10 pcs/Lot GITD Skeleton Skull ,Knife/ Flashli...	Arts, Crafts & Sewing	Skull Beads N02.	7107269291	
4	Pinkie Tm girl flower Handmade soap silicone m...	Amazon Home		7121277158	$35.00

Figure 2: First 5 records of metadata.

3 Task

At the beginning we need to find features and their identifiers such as verbs, adjectives and adverbs. The tagging process can be performed by using Stanford CoreNLP dependency parser. Through the instrument of neural networks, we' ll link entities and their descriptors and take negations into account. One step before result, entities and descriptors are produced. As of a last step, positive and negative first m feature will be showed with its first n descriptors.

The whole process is called as "Aspect based sentiment analysis".

4 Expected Results, Previous Studies

When the model run properly, for each product we' ll get good and bad features and present how many times they occur in the review for products. The output below references the related previous study. [2]

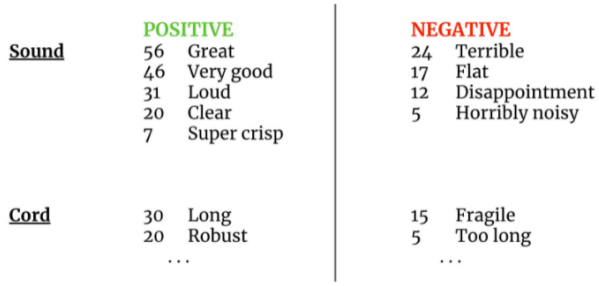

Figure 3: The expected result.

5 Potential Relevance for Master Thesis

To improve and compare the results another methods can be developed, however this step is not clear and need more research about previous studies. [3]

References

[1] J. Ni, "Amazon review data." `http://deepyeti.ucsd.edu/jianmo/amazon/`, 2018 (accessed June 12, 2021).

[2] O. Handmark, "Nlp: Making sense of review data using relation extraction." `https://towardsdatascience.com/nlp-deep-learning-for-relation-extraction-9c5d13110afa`, Jun 25,2020 (accessed June 13, 2021).

[3] E. K. A. v. d. B. Florian Kunneman, Sander Wubben, "Aspect-based summarization of pros and cons in unstructured product reviews." `https://www.aclweb.org/anthology/C18-1188.pdf`, 2018 (accessed June 13, 2021).