

# Adjective Roles in Estimating a Product Rating in a Product Review

Patrawadee Tanawongsuwan

Graduate School of Applied Statistics, National Institute of Development Administration (NIDA), Bangkok, Thailand

**Abstract**—This research work explores consumers' sentiments through written product reviews. Given a piece of product review, a person likely can tell how much the reviewer likes or dislikes the associated product. This likeness can translate into a numeric scale. For example, reading a product review, one can probably guess how many stars (which is a rating, such as, from 1 through 5) are given by the reviewer to the product. Out of hundreds of words in a particular product review, there might be only few key words that are the key estimators of the reviewer's likeness. This research work explores those estimator type of words, particularly adjectives and other non-adjectives acting as an adjective in its context. By nature, adjectives are words that describe, qualify, modify and give more information about other words. This fact hints an idea of using adjectives in a product review as an estimator for reviewer's likeness. Besides those words defined as adjectives by linguistic definition, this work is also interested in other words that might not strictly be adjectives by linguistic definition but play adjective's role in its corresponding context. This paper reports findings from exploration of the relation between adjectives in a product review and its likeness rating. Experiments are performed on a number of product reviews, book reviews to be specific, from an online bookstore.

**Keywords**—adjectives; text mining; sentiment analysis; natural language processing

## I. INTRODUCTION

Customers interested in buying a product often find out more information about the product by reading product reviews. Nowadays the internet exists almost anywhere, making almost anyone become an experience sharer. Product reviews that used to come only from critics, media, marketing team, or other professionals, can now come from people living in any corner of the world. The more sharers, the larger variety of written reviews, whether it be in terms of coverage, depth, format, formality, vocabulary, writing styles, etc. This complexity makes product reviews now a challenging problem in text mining.

Looking at a product review containing tens to hundreds or even thousands of words, there might be only few key words that represent the reviewer's points. These key words can tell how much the reviewer likes the product. Out of all kinds of words written in a review, adjectives are words that describe, qualify, modify and give more information about other words. This research work hypothesizes that adjectives in a product review can suggest the reviewer's likeness towards the product. More specifically, it studies how adjectives in a review can help estimate the review's star rating. Not only does this work

cover words that are adjectives by linguistic definition, but it also extends to those words that are not strictly adjectives but play adjective roles in their context. This extended set of adjectives will be explained further in a later part of this paper.

This work's hypothesis is tested on data from amazon.com's book section. A number of book reviews are extracted, experimented, and results reported.

Figure I shows an example of a book review, which consists of several components, such as, book title, reviewer's name, date, review's title and content, and a star rating of 1-5.

## II. RELATED WORKS

Reference [1] studies the relation between a product review's rating and different parts of speech in the review's content. Most influenced by the work, this paper extends the original idea by dealing specifically with adjectives.

As for other related works, [2] proposes a framework for analyzing and comparing consumer opinions of competing products. The proposed system, Opinion Observer, enables user to be able to see, with only a single glance of its visualization, the strengths and weaknesses of each product in the minds of consumers in terms of various product features.

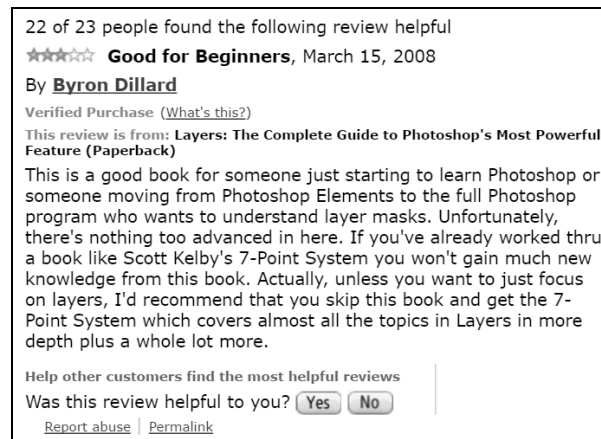


FIGURE I. AN EXAMPLE OF A BOOK REVIEW FROM AMAZON.COM. COMPONENTS ARE, FROM TOP, REVIEW'S HELPFULNESS RATING, STAR RATING, REVIEW'S TITLE, DATE, REVIEWER'S NAME, PURCHASE VERIFICATION, BOOK TITLE, REVIEW CONTENT, HELPFULNESS RATING, ABUSE REPORT, AND PERMALINK

Also, a technique based on language pattern mining is proposed to extract product features from pros and cons in a particular type of reviews.

Reference [3] presents an unsupervised learning algorithm for the classifying a reviews as being either positive or negative. The classification mechanism is based on the average semantic orientation of the phrases in the review that contain adjectives or adverbs.

References [4] and [5] study product review's sentiments in terms of positivity and negativity.

References [6] and [7] are also interested positive/negative classification of a product review.

Similarly, [8]'s work is on product review's sentiments. Proposed is a prototype system for mining topics and sentiment orientation jointly from free text customer feedback. It further demonstrates its idea in a database of car reviews.

Reference [9] introduces a system that evaluates a book review's helpfulness, using different parts of speech occurring in the review. With a similar goal though a different approach, [10] performs a detailed analysis of the review's content in order to achieve its goal.

Reference [11] looks closely on methods that seek to address the challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis. Its discussion includes topics, such as, summarization of evaluative text and broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to.

References [12] and [13] report work in opinion and blog mining. Investigating existing technology and their utility for sentiment classification on web log posts, [13] experiments on a Naive Bayes classifier in the classification of a posting's political category.

Other work with more recent updates can be found in [14], [15], [16], [17], [18] and [19].

### III. ADJECTIVES AND OTHER PARTS OF SPEECH

In a statement, every word has its role, or part of speech (POS), such as, noun, verb, adjective, and can be assigned a POS tag accordingly. The widely used Penn Treebank POS tag set [20] consists of 48 fine-grained tags, such as, NN (Noun, singular or mass), NNS (Noun, plural), NNP (Proper noun, singular), NNPS (Proper noun, plural), etc. According to the tag set, only 3 out of 48 are adjectives, namely, JJ (Adjective), JJR (Adjective, comparative), JJS (Adjective, superlative).

On the other hand, [21] lists 6 kinds of adjectives according to their roles in a statement. In other words, a word tagged as non-adjectives by the Penn Treebank approach, might still qualify as an adjective depending on its context. The combination of these two approaches broadens the adjective set to include 6 other tags as well, namely, DT (determiner), CD (cardinal number), WDT (Wh-determiner), WP (Wh-pronoun), WP\$ (Possessive wh-pronoun), and PP\$ (Possessive pronoun). This set of 9 POS tags will be called an *extended adjective set* in this paper.

## IV. DATASETS

### A. About the Dataset

As one of the most popular online stores, amazon.com carries millions of sale items in over ten departments. Listed on its website is a variety of products, each of which comes with related information, including product reviews offered by any website member. A product review might look like one in Figure I. Its components are, from top, review's helpfulness rating, star rating, review's title, date, reviewer's name, purchase verification, book title, review content, helpfulness rating, abuse report, and permalink. The star rating indicates the reviewer's overall experience towards the product. Star ratings from all reviewers are averaged and become the product's star rating. Figure II shows the average star rating, the total number of reviews, and a bar chart showing the percentages of the reviews receiving each particular number of stars.

Data for testing the hypothesis in the work are product reviews from amazon.com's Book Department during the year 2008-2016. The website amazon.com does not require book reviews to be in a certain format. Their lengths can vary from less than ten to over a thousand words. Layouts can be paragraphs, bullets, numbered lists, multiple one-line sentences, etc. It is not rare to see typographical errors, all uppercase letters, incomplete sentences, misused punctuations, non-English words, reviewer names and locations, ambiguous abbreviations, among a great variety of other things.

### B. Dataset Preparation

A book review, such as in Figure II, is processed by extracting the star rating, which is an integer between 1 through 5, inclusive. Also taken from the review are review's title and textual content, which then go through the process of POS tagging. This paper adopts the well-known Stanford Log-linear Part-Of-Speech Tagger [22], [23]. The process of POS tagging is much more complex than simply labeling each word according to a predefined list of word-tag pairs. The tagger looks at each word as well as its context in order to determine its part of speech. Figure III shows the same product review from Figure I after being POS-tagged.



FIGURE II. LISTED WITH EACH PRODUCT ARE THE AVERAGE STAR RATING, THE TOTAL NUMBER OF REVIEWS, AND A BAR CHART SHOWING THE PERCENTAGES OF THE REVIEWS RECEIVING EACH PARTICULAR NUMBER OF STARS

POS tags in the Penn Treebank set are very particular. For instance, there are six tags for slightly different types of verb, four for noun, three for adjective, etc. In this study, similar

tags are group together to become a set of *aggregated tags*, which are used in addition to the original tags.

After a review gets POS-tagged, the review content is summarized by the number of times each tag appears there in. Its corresponding percentage is then calculated. As for the review title, similar numbers are obtained, with a slight difference, however. Since a review title usually contains only few words, counting the number of times each tag occurs might not be quite meaningful. Instead, each tag is observed in terms of its occurrence (or lack thereof) in the review title, regardless of the number of times it might repeat. All of these quantities are noted and will be used as input in the next step.

## V. METHODOLOGY

Recalling that this work's objective is to study how adjectives in a product review play roles in estimating the review's star rating. A set of experiments are set up just for such purpose.

### A. Experiment 1: All Tags on Neural Networks

In this base experiment, the neural network classifier [24] is chosen as a learning model because it is known to handle continuous and numeric values fairly well. Input to the model are the number and percentage of occurrences of each tag, both original and aggregated tags, in both review title and content. Be reminded that there are 48 original and 8 aggregated tags. Output is the star rating of the review.

Good\_NNP for\_IN Beginners\_NNPS This\_DT is\_VBZ a\_DT good\_JJ book\_NN for\_IN someone\_NN just\_RB starting\_VBG to\_TO learn\_VB Photoshop\_NNP or\_CC someone\_NN moving\_VBG from\_IN Photoshop\_NNP Elements\_NNP to\_TO the\_DT full\_JJ Photoshop\_NNP program\_NN who\_WP wants\_VBZ to\_TO understand\_VB layer\_NN masks\_NNS .\_. Unfortunately\_RB .\_. there\_EX 's\_VBZ nothing\_NN too\_RB advanced\_JJ in\_IN here\_RB .\_. If\_IN you\_PRP 've\_VBP already\_RB worked\_VBN thru\_NN a\_DT book\_NN like\_IN Scott\_NNP Kelby\_NNP 's\_POS 7-Point\_JJ System\_NN you\_PRP wo\_MD n't\_RB gain\_VB much\_JJ new\_JJ knowledge\_NN from\_IN this\_DT book\_NN .\_. Actually\_RB .\_. unless\_IN you\_PRP want\_VBP to\_TO just\_RB focus\_VB on\_IN layers\_NNS .\_. I\_PRP 'd\_MD recommend\_VB that\_IN you\_PRP skip\_VBP this\_DT book\_NN and\_CC get\_VB the\_DT 7-Point\_JJ System\_NN which\_WDT covers\_VBZ almost\_RB all\_PDT the\_DT topics\_NNS in\_IN Layers\_NNS in\_IN more\_JJR depth\_NN plus\_IN a\_DT whole\_JJ lot\_NN more\_JJR .\_.

Tags: CC=Coordinating conjunction, DT=Determiner, EX=Existential there, IN=Preposition/subordinating conjunction, JJ=Adjective, JJR=Adjective (comparative), MD=Modal, NN=Noun (singular or mass), NNS=Noun (plural), NNP=Proper noun (singular), NNPS=Proper noun (plural), PDT=Predeterminer, POS=Possessive ending, PRP=Personal pronoun, RB=Adverb, TO=to, VB=Verb (base form), VBG=Verb (gerund/present participle), VBN=Verb (past participle), VBP=Verb (non-3<sup>rd</sup> person singular present), VBZ=Verb (3<sup>rd</sup> person singular present), WDT=wh-determiner, WP=wh-pronoun

FIGURE III. A REVIEW'S TITLE AND TEXTUAL CONTENT FROM FIGURE I AFTER BEING POS TAGGED

TABLE I. AGGREGATED TAGS AND THEIR MEMBERS

Aggregated Tags	Members
Noun	noun (singular or mass) noun (plural) proper noun (singular) proper noun (plural)
Proper Noun	proper noun (singular) proper noun (plural)
Pronoun	personal pronoun possessive pronoun
Verb	verb (base form) verb (past tense) verb (gerund/present participle) verb (past participle) verb (non-3 <sup>rd</sup> person, singular, present) verb (3 <sup>rd</sup> person, singular, present)
Adjective	adjective adjective (comparative) adjective (superlative)
Adverb	adverb adverb (comparative) adverb (superlative)
Wh	wh-determiner wh-pronoun possessive wh-pronoun wh-adverb
Punctuation	# \$ . , : ( ) " ' " ' "

### B. Experiment 2: All Tags on Bayesian Network and Decision Tree

This experiment is similar to experiment 1 except that the Bayesian Network [25] and Decision Tree [25] are used as learning models. Since these two models handle discrete values, they are chosen with the hope that they might bring different perspectives from those given by neural networks. Experiment 2 is also considered another base experiment since all tags are used.

### C. Experiment 3: Non-Adjectives Tags on Neural Networks

In this experiment, all adjectives, whether they be those tagged as adjectives according the Penn Treebank tag set or the extended adjective set defined in this paper, are eliminated from the input set. The neural network is chosen as a classification model. Results are expected to show that the lack of adjectives can negatively affect the star rating estimation.

### D. Experiment 4: Non-Adjectives Tags on Bayesian Network and Decision Tree

Experiment 4 is a combination of previous experiments. It combines the learning models from experiment 2 with the input from experiment3. It might be interesting to see how these discrete models perform with the reduced set of input.

## VI. EXPERIMENTAL RESULTS AND DISCUSSION

Results from experiments 2 and 4, reported in Table III, come out better than those from experiments 1 and 3. This is quite surprising since the neural network classifier, which is expected to handle continuous and numeral data quite well, does less well compared with Bayesian Network and Decision Tree.

According to Table II, the estimation accuracies when adjectives are used (experiment 2) are better than those when adjectives are left out (experiment 4). This is true for both training and test sets, across both classification models. This observation validates this study's hypothesis, saying that adjectives in a review can be used as an estimator for the star rating. However, declaring such a statement needs to be careful because it makes a claim about adjectives but not saying that adjectives are the best or even better than other POS. Also shown in Table II are the mean absolute errors, which are consistently higher when adjectives are left out (experiment 4) than those when adjectives are used (experiment 2).

As for other observations, the estimation accuracies are significantly lower for the test set than training set, consistently for both models. The mean absolute errors are in an acceptable range.

TABLE II. CLASSIFICATION PERFORMANCE BY BAYESIAN NETWORK AND DECISION TREE

	Experiment 2		Experiment 4	
	Accuracy	Mean Absolute Error	Accuracy	Mean Absolute Error
<b>Bayesian Network</b>				
Training Set	58.16%	0.2132	50.57%	0.4487
Test Set	49.44%	0.2491	39.61%	0.5013
<b>Decision Tree</b>				
Training Set	83.67%	0.0905	79.51%	0.2140
Test Set	48.82%	0.2084	41.08%	0.3867

## VII. CONCLUSION

This research studies product reviews and focuses on the role of adjectives. It is hypothesized that adjectives in a product review can help estimate the review's star rating. A set of experiments are designed and carried out on a set of book review data from amazon.com. Experimental results agree with the hypothesis. For future work, this study can be extended to cover more depth and discover more insights.

## REFERENCES

- [1] P. Tanawongsuwan, "Relation Between a Book Review Content and Its Rating," in *Proc. International Conference on Computer Information Systems and Industrial Applications*, Bangkok, 2015, pp. 853-856.
- [2] B. Liu, M. Hu and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in *Proc. 14<sup>th</sup> Int'l Conf. on World Wide Web (WWW 2005)*, Chiba, Japan, May 2005.
- [3] P. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," in *Proc. 40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, Philadelphia, Pennsylvania, USA, Jul. 2002.
- [4] P. Tanawongsuwan, "Product review sentiment classification using parts of speech: A case study of textbook reviews," in *Proc. 3<sup>rd</sup> IEEE International Conference on Computer Science and Information Technology*, Chengdu China, 2010, pp. 424-427.
- [5] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proc. of Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, 2002.
- [6] H. Cui, V. Mittal and M. Datar, "Comparative experiments on sentiment classification for online product reviews," in *Proc. 21st National Conference on Artificial Intelligence (AAAI 2006)*, Boston, Massachusetts, USA, Jul. 2006.
- [7] K. Dave, S. Lawrence and D. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," in *Proc. 12th Int'l Conf. on World Wide Web (WWW 2003)*, Budapest, Hungary, May 2003.
- [8] M. Gamon, A. Aue, S. Corston-Oliver and E. Ringger, "Pulse: mining customer opinions from free text," *Proc. International Symposium on Intelligent Data Analysis (IDA 2005)*, Sep. 2005, Madrid, Spain.
- [9] P. Tanawongsuwan, "Part-of-speech approach to evaluation of textbook reviews," in *Proc. International Conference on Industrial and Intelligent Information*, Bangkok, 2010, pp. 352-356.
- [10] P. Tanawongsuwan, "Content-based evaluation of textbook reviews," in *Proc. 6<sup>th</sup> International Joint Conference on Computer Science and Software Engineering*, Phuket Thailand, 2009.
- [11] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, 2.1-2 (2008): 1-135.
- [12] G. Attardi and M. Simi, "Blog mining through opinionated words," in *Proc. 15th Text Retrieval Conference (TREC 2006)*, Gaithersburg, Maryland, USA, Nov. 2006.
- [13] K. Durant and M. Smith, "Mining sentiment classification from political web logs," in *Proc. KDD Workshop on Web Mining and Web Usage Analysis (WebKDD 2006) in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006.
- [14] H. Khatter and B.M. Kalra, "A new approach to blog information searching and curating," in *Proc. CSI 6<sup>th</sup> Int'l Conf. on Software Engineering (CONSEG)*, vol., no., pp.1.6, 5-7 Sept. 2012, doi:http://dx.doi.org/10.1109/CONSEG.2012.6349518
- [15] S. Zerr, N. K. Tran, K. Bischoff, and C. Niederee, "Sentiment analysis and opinion mining in collections of qualitative data," in *Proc. 10<sup>th</sup> Int'l Conf. on Preservation of Digital Objects*, Lisbon, Portugal, 2-6 September 2013.
- [16] T. Mumu, "Social Network Opinion and posts mining for community preference discovery," *Masters Thesis*, School of Comp. Sci., U. of Windsor, Ontario, Canada, 2013.
- [17] S. Malarvizhi, S. Prasana and A. Muthukumaravel, "Opinion ana data mining of customer feedback data on the web," *Int'l Jour. of Latest Trends in Engineering and Technology (IJLTET)*, vol. 2, issue 4, pp. 222-225, July 2013.
- [18] K. S. M. Anbananthen and A. M. H. Elyasir, "Evolution of opinion mining," *Australian Jour. of Basic and Applied Sciences*, vol. 7, no. 6, pp. 359-370, April 2013.
- [19] G. W. Sathianesan and S. Sankaranarayanan, "Unified semantic blog mining framework and summarized blog retrieval," *Jour. of Comp. Sci.*, vol. 9, no. 2, pp. 207-217, April 2013, doi:10.3844/jcssp.2013.207.217.
- [20] M. P. Marcus, B. Santorini and M. A. Marcinkiewicz, "Building a large annotated corpus of English: The Penn Treebank," *Computational Linguistics*, vol. 19, no. 2, pp. 313-330, 1993.
- [21] A. J. Thomson and A. V. Martinet, *A Practical English Grammar*, 4<sup>th</sup> ed., Oxford University Press, 1986, ch. 3.
- [22] K. Toutanova and C. D. Manning, "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger," in *Proc. of the Joint SIGDAT Conf. on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*, 2000, pp. 63-70.
- [23] K. Toutanova, D. Klein, C. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proc. of HLT-NAACL*, 2003, pp. 252-259.
- [24] C. Stergiou and D. Siganos, "Neural networks," *Computer Science Dept. Univ. of U.K. Journal*, vol. 4, 1996.
- [25] I. H. Witten, E. Frank and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3<sup>rd</sup> ed., Morgan Kaufmann, 2011, ch. 6.