

A Comparative Study between Factor Based Sentiment and Overall Sentiment

Santanu Modak^{1*}, Abhoy Chand Mondal²

^{1,2}Department of Computer Science, University of Burdwan

Corresponding author: modaksantanu@gmail.com, Tel.: +91-9153013838

Available online at: www.ijcseonline.org

Abstract— Sentiment Analysis has pulled in significantly more consideration from analysts in late years. As web based shopping is getting to be typical, more item data and item audits are posted on the Internet. Since clients can't see and feel the items straightforwardly, item surveys are turning into a basic wellspring of subjective data. Accordingly, the volume of audits is expanding drastically. It is hard to a client to peruse every one of the surveys of related item and contrast and other item in view of audits. Some of the time there is a contrast between in general assessment of the item and supposition about each feature of a similar item. In this paper, we examine 480 smart phone surveys from famous online business site and endeavor to locate a similar contrast. We allot fuzzy score for each sentiment word and figure arithmetic mean of the allocated fuzzy scores. Examination results demonstrate that connection between the general assessment and aftereffect of feature extraction undertaking, and the promising execution of our methodology has likewise been appeared.

Keywords—Feature Based Sentiment Analysis, Fuzzy, Sentiment Phase Detection, Sentiment Dictionary

I. INTRODUCTION

Sentiment analysis is a type of text analysis, under the broad area of natural language processing, computational linguistics and text mining [1], that examinations presumption in a given scholarly unit with the objective of appreciation the polarities of the suppositions imparted and the sorts of emotions toward various parts of a review. With the striking progression of Web 2.0 in the most recent decade, correspondence stages, such as web journals, wikis, online discussions, and long range interpersonal communication gatherings, have turned into a rich information digging hotspot for the location of open opinions. As the item data is expanding in e-commercial center, numerous item audits are likewise posted in the Internet. Since clients can't see the items straightforwardly, item surveys are turning into a fundamental wellspring of subjective data.

With developing needs of item audits, the issue of the monstrous volume of surveys emerges as a commonsense issue. It is important to gather and investigate the Shopper opinions. Therefore, we propose Factor based opinion investigation for preparing feeling data from e-commerce sites. Our quantifiable examination of this data offers verification to the nearness of addressing reviews, and furthermore for the simultaneousness of positive and negative emotions inside overviews. We find a strong

inclination towards broad characteristics in the assertion of positive sentiments, while negative ones are simply more fairly circled. We recognized different time components of the generation of studies subject to the closeness of promoting and verbal effects.

Notwithstanding staggering undertakings were made around there, the present Sentiment Analysis systems can't give understanding into purchaser's lead. Simply more especially, a customer may give low assessing on a feature yet in the meantime get it as they like it's another feature. The undeniable affirmation of the key highlights, which on a to a great degree major level in familiar customer's essential technique can give business authorities course on what features to upgrade and how to move the offer of their things. Tragically, the present component based thought layout structures are simply prepared to give summary of presumption from overviews, yet don't allow to effectively survey the essentialness of every segment in association with customer's tendency. So, we analyze 480 Smartphone reviews from popular online business website and attempt to find a comparative complexity. We assign fuzzy score for every assessment expression of six chose features of Smartphone and figure number arithmetic mean of the designated fuzzy scores. Our outcomes exhibit that association between the general appraisal and eventual outcome of feature extraction undertaking.

II. RELATED WORK

The exploration of Feature Based Sentiment mining was initially planned by Hu [2], and is made out of the accompanying four subtasks: 1) Identifying item includes from survey remarks; 2) Identifying sentiment words with respect to item includes; 3) Determining the extremity of feeling words; 4) Polarity assurance of audit with respect to item includes.

In [3], scientists propose an all encompassing way to deal with product review summarization. There are 3 stages in a product review summarization: include feature extraction, sentiment analysis, and feature scoring. While extracting product features and sentiment words, scientists present an adaptable extraction instrument, which can be redone as indicated by the attributes of source reports. For sentiment analysis and feature scoring, specialists use rate scores, which are clients' general evaluations of the item. Feature dependent polarities of sentiment words are additionally used. In conclusion, scientists demonstrate that this contemplation can enhance the precision of item audit rundown with tests.

In [4], analysts present another dataset of item surveys from Amazon.com, with passionate data removed by slant identification apparatuses. Their factual investigation of this information gives confirmation to the presence of questioning surveys, and in addition for the conjunction of positive and negative feelings inside audits. Analysts locate a solid predisposition towards huge qualities in the statement of positive feelings, while negative ones are all the more uniformly dispersed. Analysts recognized diverse time elements of the making of surveys reliant on the nearness of advertising and informal impacts. Analysts characterize an operator based model of the clients of item survey networks utilizing a displaying system for online feelings. That model can recreate the situations of reaction to outer impacts, and a few properties of the dissemination of positive and negative feelings communicated in item surveys.

In [5], Researchers introduced another method into feature based sentiment summary framework which can assess the significance of highlights in item (hotel) rating procedure of purchasers by the utilization of a few measurements in fuzzy framework, for example, Shapley esteem and Interaction Index, and assess their viability in conclusion mining use of hotel industry. In [6], Researchers introduced a structure whose focuses are twofold: (I) the utilization of a learning approach prepared to indicate fuzzy functions used for building the associations graph addressing the reasonableness between sentiment concepts and different domains; and (ii) the headway of a semantic resource in light of the relationship between an extended type of WordNet, SenticNet, and ConceptNet, that has been used both for

removing thoughts and for classifying sentences inside specific spaces. In [7], Researchers proposed a technique based on quantified sentiment lexicon and fuzzy set, constructed some evaluated estimation dictionaries in light of three Chinese notion vocabularies, and after that compute notion force of Chinese audits by measured notion dictionary, at last, Researchers group Chinese surveys in view of fuzzy classifier.

III. METHODOLOGY

The components of our review mining system can be sketched out with essential subtasks:

Step 1: Collection of feedback: It is more calm and successful than general society review to accumulate the buyer notion from the online business structure.

Step 2: Distinguish product features: center around the extraction of express product features, recognizing comparing client assessments about these product features and deciding their extremity.

Step 3: Assessment: The computational generation of the framework can be utilized to assess the extremity of sentiment words (essentially thought about descriptive words and intensifiers) of each component. The goal was not to decide suppositions or the creator's interior state, notwithstanding, yet to recognize the job of communicated feeling for online correspondence. Consequently the focal point of the undertaking was to recognize the opinion communicated in each message, in the case of mirroring the writer's shrouded inward express, the expected message understanding, or the per user's concealed inside state. The extremity of words is recovered from the bundle design and the review extremity is computed utilizing: Sum of extremity of the considerable number of sentiment words in a review separated by the aggregate number of words in the review.

Step 4: Compute Percent Error: When you figure results that are going for known qualities, the percent error equation is valuable instrument for deciding the exactness of your estimations.

$$\text{Percent Error} := \left| \frac{\text{Hypothetical Value} - \text{Experimental Value}}{\text{Hypothetical Value}} \right| \times 100$$

The experimental value is your figured esteem, and the hypothetical value is your known esteem. A rate near zero means you are near your focused on esteem, which is great.

Step 5: Compute Pearson's correlation coefficient to demonstrate the connection between the general assessment and aftereffect of feature extraction undertaking, and the promising execution of our methodology has likewise been appeared.

:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where x = hypothetical value, y = experimental value, n = number of sample reviews.

IV. ANALYSIS

Collection of feedback:

The dataset for our algorithm contains messages in numerous assumptions on solitary spaces, Smartphone, coming from internet business site. We have gathered a gigantic dataset of 480 surveys from Amazon.in to investigation the elements of feelings identified with products. For each sentence in an audit, we commented on the item angles, and the comparing notion words related with the viewpoint together with the polarity (fuzzy score) of the opinion words. Some notion words comparing to every one of the perspectives in the sentence, so they are explained as all. In view of the rundown of conclusion words, we see that there are six angles.

SAMPLE DATASET

Table 1

Phone Name	Back camera	Font camera	Colour/looking	Battery	Charging time	Screen&Display	Ratings
VIVO V7 (Matte Black, 32 GB) (4 GB RAM)	Awesome	Awesome	Troubo				5
	good	good	Very classy			Split	5
	good	good		good			5
	little bit	superb	nice	Good			5
	not good	not good					1
				not good			1
	Great	Great	Great	great	nearly 1.5-2 hours	not fare	5
	Excellent	Excellent					5
	good	good		good		good	5
					best		3
OnePlus 5T (Midnight Black 6GB RAM + 64GB memory)	Better			A day average		Excellent	2
	Very nice			one day	Quickly		5
	Flash	Flash				m	5
	Nice	Nice				Beautiful	5
	Good	Good	Good	Good	Good	Good	5
	Good	Good		Good		Excellent	5
	Great	Great	Sleek	1 day	Amzing,20mins		5
					Good	Good	4
	Good	Good	Good				5
					Quick	Wonderful	4
					30 mins only	Awesome	2
			Killer		Good		5
			Disappoint				5
	Best	Best				Awesome	5
	Best	Best					5
					Excellent		5
	Excellence	Excellence		un-believable	70% in barely 20-25 mins		5
	good	good				Large	4
	Awesome	Awesome	Very slim		Fast	good	5
	Very good	very good		Amazing		Very good	4
				half & hour	Lovely	5	
Good	Good	Bright Sunlight			amazing	5	
Loving				Less time		5	
didn't like				Dash	didn't like	3	
		beautiful		Dash	bright	4	
		beast				2	
Sony Xperia XA1 (Gold, 32 GB) (3 GB RAM)	Best of best	best	best				5
	awesome		WoW				5
	Very Good	Very Good					5
						problem waist	1
					problem		1
	good	good	Nice	despite		Neat	5
	nice	nice	decent				4
	shutdown	shutdown					1
	awesome	awesome	Beautiful	works like		better	5
	excellent	excellent				better	4
very low	very low	useless				1	
			not enough			1	
best	best		very good	full day	better	4	
awesome	Excellent			Quick	Superb	5	
good	good	good			great	4	

DISTINGUISH PRODUCT FEATURES:

We identified six different features of smartphone , namely, Back Camera, Front Camera, Color/Looking, Battery, Charging Time, Screen & Display.

Features Occurrences in the reviews

Table 2

Feature	Number of occurrences in the review	Percentage of occurrences $\frac{\text{Number of Occurrence}}{\text{Total no. of Reviews}} \times 100$
Back Camera	388	80.83%
Font Camera	362	75.41%
Looking	237	49.37%
Battery Life	303	63.12%
Charging Time	183	38.12%
Screen and Display	258	53.75%

Sample of Feature specific fuzzy score allocation Back Camera

Table 3

Sentiment Word / Phases	Assigned Fuzzy Score
Best of Best, Very Nice, Best	5
Decent, Good enough	4
Better, Good clarity	3
Below average, not so good, low light, below expectation	2
Worst, very very bad	1

Front Camera

Table 4

Sentiment Word / Phases	Assigned Fuzzy Score
Best, awesome	5
Very good, performer	4
Okey, better	3
Disappointed, poor, not great	2
Very low, bad, pathetic	1

Looking

Table 5

Sentiment Word / Phases	Assigned Fuzzy Score
Awesome, very classy, killer, great, beautiful, stunning	5
Very slim, sleek, wow, nice, premium	4
Good, elegant, okey	3
useless	2
Very poor	1

Compute Percentage Error:

the percent error equation is valuable instrument for deciding the exactness of estimations.

$$\text{Percent Error} := \left| \frac{\text{Hypothetical Value} - \text{Experimental Value}}{\text{Hypothetical Value}} \right| \times 100$$

The experimental value is your figured esteem, and the hypothetical value is your known esteem.

Sample Percentage Error:

Table 6

Ratings	ANALYSIS	Arithmetic Mean	PERCENTAGE OF ERRORS
5	(5+5+2)/3	4	20
5	(4+4+5+2)/4	3.75	25
5	2	3.666666667	26.66666667
5	(3+5+5+3)/4	4	20
1	(1+1)/2	1	0
1	(1)/1	1	0
5	(5+5+5+5+1)/5	4.2	16
5	(5+5+3)/3	4.333333333	13.33333333
5	(4+4+3+3)/4	3.5	30
5	(5+3+5)/3	4.333333333	13.33333333
5	(3+3+5)/3	3.666666667	26.66666667
5	(4+5+5)/3	4.666666667	6.666666667
5	(4+4+3+3+3)/6	3.333333333	33.33333333
5	(4+4+3+5)/4	4	20
5	(5+5+3+3+5)/5	4.2	16
4	(3+3)/2	3	20
5	(4+4+3)/3	3.666666667	26.66666667
5	(2+3)/2	2.5	50
5	(4+5+5)/3	4.666666667	6.666666667
5	(4+5)/2	4.5	10
5	(5)/1	5	0
5	(5+5+1+3)/4	3.5	30
4	(4+4+4)/3	4	0
5	(5+5+4+4+3)/5	4.2	16
5	(3+5)/2	4	20
5	(4+4+5+5)/4	4.5	10
5	(5+4)/2	4.5	10
3	(1+2+1)/3	1.333333333	33.33333333
4	(5+2+4)/3	3.666666667	6.666666667
5	(5+5+5)/3	5	0
5	(5+4)/2	4.5	10
5	(5+5)/2	5	0
1	(1)/1	1	0
1	(1)/1	1	0
5	(4+4+5+2+2)/5	3.4	32

Compute Pearson’s correlation coefficient :

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where x = hypothetical value, y = experimental value, n = number of sample reviews. The connection between the general assessment and aftereffect of feature extraction undertaking is 0.7599.

V. CONCLUSION and FUTURE SCOPE

In this paper, we made a comparative study between Factor based Opinion and Overall Opinion of a product (e.g. Mobile Phone). Firstly, we collect actual ratings and complete review from e-commerce website. Then we separate factor based opinion word/ phase. Then we made a sentiment word/phase dictionary of every factor. A score allotted to every word/phase of the dictionary. Then we calculate arithmetic mean of the factor based score; calculate percentage error and correlation coefficient. In future we will expand our research to find sarcastic opinion study.

REFERENCES

- [1] J. Wang (ed), *Encyclopedia of Data Warehousing and Mining* (Information Science Reference, Hershey, 2008).
- [2] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Seattle, WA, USA, August 2004, pp. 168–177.
- [3] J. Yang, J. Myung and S. Lee, "A Holistic Approach to Product Review Summarization," *2009 Software Technologies for Future Dependable Distributed Systems*, Tokyo, 2009, pp. 150-154. doi: 10.1109/STFSSD.2009.
- [4] D. Garcia and F. Schweitzer, "Emotions in Product Reviews--Empirics and Models," *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, Boston, MA, 2011, pp. 483-488. doi: 10.1109/PASSAT/SocialCom.2011.
- [5] H. Q. Vu, G. Li and G. Beliakov, "A fuzzy decision support method for customer preferences analysis based on Choquet Integral," *2012 IEEE International Conference on Fuzzy Systems*, Brisbane, QLD, 2012, pp. 1-8. doi: 10.1109/FUZZ-IEEE.2012.6250776
- [6] Dragoni, Mauro, Andrea GB Tettamanzi, and Célia da Costa Pereira. "A fuzzy system for concept-level sentiment analysis." *Semantic web evaluation challenge*. Springer, Cham, 2014.
- [7] Kamal, A.; Abulaish, M., "Statistical Features Identification for Sentiment Analysis Using Machine Learning Techniques," *Computational and Business Intelligence (ISCBI), 2013 International Symposium on*, vol., no., pp.178,181, 24-26
- [8] Popescu, Ana-Maria, and Orena Etzioni. "Extracting product features and opinions from reviews." *Natural language processing and text mining*. Springer, London, 2007. 9-28.

- [9] Jefferson, Chris, Han Liu, and Mihaela Cocea. "Fuzzy approach for sentiment analysis." *Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on*. IEEE, 2017.
- [10] Lek, Hsiang Hui, and Danny CC Poo. "Aspect-based Twitter sentiment classification." *Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on*. IEEE, 2013.
- [11] Lalithamani, R. A. N., Leela Sravanthi Thati, and Rakesh Adhikesavan. "Sentence level sentiment polarity calculation for customer reviews by considering complex sentential structures." *IJRET: International Journal of Research in Engineering and Technology* 3 (2014).
- [12] Ahlgren, Per, Bo Jarneving, and Ronald Rousseau. "Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient." *Journal of the American Society for Information Science and Technology* 54.6 (2003): 550-560.

Authors Profile

Santanu Modak is Research Fellow in the Department of Computer Science, University of Burdwan. He received his B.Sc(Hons) and M.Sc in Computer Science degrees in 2009 and 2011 respectively. He qualified UGC NET three times in Computer Science and Applications.



He published four international journals. He is a life member of Indian Science Congress Association (ISCA), member of International Association of Computer Science and Information Technology (IACSIT) and member of International Association of Engineers.

Abhoy Chand Mondal is currently Professor of Department of Computer Science, Burdwan University, W.B., India. He received his B.Sc.(Mathematics Hons.) from The University of Burdwan in 1987, M.Sc. (Math) and M.C.A. from Jadavpur University, in 1989, 1992 respectively. He received his Ph.D. from Burdwan University in 2004. He has 1 year industry experience and 18 years of teaching and research experience. No. of papers more than 50 and no of journal published is 25.

