International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

A New Feature based Scoring Technique to Discover Sentiments Present in the Online Product

R.Bhuvaneswari, S.Ravichandran

Abstract: ***The increasing usage of internet, online stores and social media has provided the users to express their opinion, attitude and views without any reluctance and fear on the World Wide Web. These opinions expressed by the users can be related to a product or service as well as any global issues. The colossal growth of the web technology has offered the consumers to know more about the products they intend to buy from the existing customer's reviews. This paper focuses on analyzing the opinion by splitting the positive and negative opinion and then guides the user about the ground truth regarding the performance and quality of the product. The important idea here is to categorize the important features of the product and then accordingly provide the feature wise computation instead of roughly promoting a product as good or bad.***

Keywords : sentiment analysis, opinion mining, data mining, knowledge discovery.

I. INTRODUCTION

Sentiment analysis is computational process of discovering the individuals' opinions, sentiments, feelings, and dispositions. According to the author [Bing Liu, 2015] this entrancing issue is progressively critical in organizations and society. Sentiment analysis is a sort of Natural Language Processing (NLP) for following the state of mind of general society about a specific item or service. Sentiment analysis, additionally called opinion mining, is area of study related to opinions, sentiments, disposition, feelings, assessments, evaluations towards substances, for example, items, services, associations, people, issues, events, subjects, and their traits. According to the author [M.R. Saleh et al., 2011] enormous volume of data related to the client Opinions or views is very cumbersome to assess and need a computerized general approach to deal with the opinions gathered in the internet. Bounteous forums, journals, online business, social media, news reports and more importantly the web based resources aids the users to express their opinions, which can be utilized for understanding the opinions of the overall population and customers on online sites,

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Mrs. Bhuvaneswari*, R Asst Professor of Computer Science, Sudharson College of arts and science Pudukkottai
Dr. S. Ravichandran, HOD and Assit.Prof, Computer Science Department , H.H.The Rajah's college, Pudukkottai.

social media, political activities and product sentiments. When a person wants to buy a product online he or she will typically start by searching for opinions written by other people on the various offerings.Hundreds of startups are developing sentiment analysis solutions and this area of research is gaining momentum of late. Sentiments and opinions are becoming more important mainly due to rapid growth of e-commerce sector, which provides copious data source for analyzing sentiments and views. Of late, most of the customers on e-commerce website depend heavily on the views and opinions posted by existing customers and users before taking the decision to buy the products. For example opinions given on e-commerce sites like Amazon, FlipKart can influence the customer's decision in buying products online.

II. DATA CHARACTERISTICS

Review related to products is an independent view or suggestion the users post in real time messages and have plethora of distinct characteristics, which throws new challenges in carrying out sentiment analysis. The following are the most important characteristics of a review message provided by the users,

A. Data Length

The data length of the different user will be unique and the length of the message has no constraint and this makes the analysis quite cumbersome as the analysis has to be carried out on different longer size opinions.

B. Writing Technique

The users usually deals with short form writing like ASAP, wrong spellings and slang related to cyber technology in their views. Mostly the user data are of different sizes, usage of acronym, and use emoticon and other special characters to convey special meanings to their views.

C.Emoticons

Emoticons are tiny pictures of facial expression using punctuation and letters and the consumer use the emoticons to express their views and moods. Every emoticon has a special mood resemblance related to the product.

D. Data Used

The data to be used in the proposed work in this paper is acquired from gsmarena.com, a portal specialized in reviewing mobile phones and comprises of huge volume of unstructured data in the form of views and remarks provided by the users.

III. OVERALL BLOCK DIAGRAM

The words present in the view are the terms user use to define their attitude and opinion (positive, negative or neutral). To discover the exact product opinion of the user, a new approach to extract the important features about the product and discover the opinions related to the features is proposed.

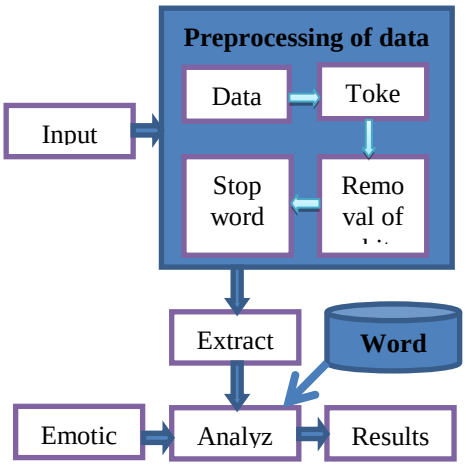

Figure 1: Overall Block Diagram

The preprocessing technique is illustrated here in this section, the first process is tokenization and this tokenization is nothing but separating words present in the view based on the common space separator. The tokenized words are stored separately with the opinion ID and the procedure to carry out the tokenization is shown in the figure 2.

The tokenized words are then processed to remove the stop words or common words which are no longer useful in determining the orientation of the opinion and words like "and", "the", "so", "that", "what" are removed from the list of words to reduce the size of the words to be processed.

Procedure TokenizeWords (Opinion Data D)

INPUTS: Opinion Data File D
OUTPUT: Tokenized words

1. Load Opinion Data D
2. While [D ≠ EOF] do begin
3. Word=ϕ
4. For each character C in D do begin
5. IF [C ≠ Space]
6. Word = Word + C
7. Else
8. Store Word in Array
9. Word = ϕ
10. End IF
11. End For
12. End While
13. Return Array

END procedure

Figure 2: Pseudo code to tokenize the words

Procedure RemoveStopWord (Tokenized Word Array TA, Stop word reference file SF)

INPUTS: Tokenized Word Array TA,
Stop words reference file SF
OUTPUT: Tokenized words without stop words

1. Load Tokenized Word Array TA
2. For each word W in TA do begin
3. IF [W present in SF]
4. Remove W
5. End IF
6. End For
7. Return Array TA

END procedure

Figure 3: Pseudo code for stop word removal

The figure 3 shows the pseudo code to remove the stop words that are no longer has any effect to the overall opinion or expression of the user. Since the young and new users use many emoticons to represent their emotions regarding the product, the emoticons are extracted and compared with the emoticon reference file to compute the emoticon value as shown in figure 4.

Figure 4: Pseudo code to find emoticon value

Procedure EmoticonValue (Tokenized Word Array TA, Emoticon reference file EF)

INPUTS: Tokenized Word Array TA, Emoticon reference file SF
OUTPUT: Emoticon value with OpinionID

1. Load Tokenized Word Array TA
2. For each word W in TA do begin
3. IF [W not present in Dictionary]
4. IF [W present in EF]
5. Compute the Emoticon value →E-value with ID
6. End IF
7. End IF
8. Return E-value with ID

END procedure

Table 1: Emoticon reference file

Emoticon Reference file		
Emoticon	**Meaning**	**Value**
;) :) =) :-)	Happy	0.5
XD	Laughing	1
:D	Big smile	1
BD	Big Smile with glass	1
`:*	kiss	0.5
:| :\	undecided	0
:(	Sad	-0.5
B(	Sad with goggles	-0.5
:'(	Crying	-1
X-(	Angry	-1

The feature extraction is carried out to extract three basic attributes from the opinions and they are Appearance,

Affordability and Essential. Here the appearance is the look and the style, sleekness, weight, colors, design, and sturdiness. The affordability deals with the price tag, tariff, and the rate of the product. The essentials are basically the hardware and the software used like battery, screen size, screen resolution, speakers, camera, connectivity and processor. The tokenized document is further segregated according to the three features explained earlier and the tokens are stored in the respective feature heads. A special feature reference document is used to separate the review tokens according to the basic three features. The pseudo code to extract the basic features is shown in the figure 5.

```
Procedure ExtractFeature ( Tokenized
Word TA, Feature reference doc FD )
INPUTS: Tokenized Word TA, Feature reference doc FD
OUTPUT: Emoticon value with OpinionID
1.  Load Tokenized Word Array  TA
2.  For each word W in TA do begin
3.  IF [ W present in AppearanceSection of FD]
4.  Fetch the complete phrase tokens
5.  Store in their Appearance array
6.  ElseIF [ W present in AffordabilitySection of
    FD]
7.   Fetch the complete phrase tokens
8.  Store in their Affordability array
9.  Else [ W present in EssentialSection of FD]
10.  Fetch the complete phrase tokens
11. Store in their Essential array
12. End IF
13. End For
END procedure
```

Figure 5: Pseudo code of basic feature extraction.

The sample special feature reference document file is showcased in the figure 6 and this file is used to segregate the important features present in the mobile phones and the sentiment analysis is carried out based on these three features extracted. Since the opinions will have a mixed responses like, a great camera capabilities, good screen resolution but will be having poor battery, high rate and very less signal strength. Here the proposed approach provides a new method by delivering the opinion values separately according to the features and this will definitely offer a better insight related to a product and the consumer will experience the real ground truth about a product with break up details and has an opportunity to know about the strengths and the weakness of the product.

Table 2: Basic Special features reference file

Special Feature Reference file	
Tokens	**Meaning**
Price, Rate, affordable	**Affordability Feature**
Tariff, Cost, $	
Expensive, Cheap	
Amount, INR, Economy	
Screen, Size, build, inch	**Appearance Feature**
Design, Sleek, Sturdy	
Weight, Grams, pixels, resolution	
Lavish, Classy, Elegant	
Battery, body, Camera	**Essential Features**
MP, FPS, Video, Face	
OS, Dual, Glass, SIM	
Hybrid, Octo, Quad,	
USB, OTG, Wifi, Charger	
NFC, Chip, RAM, External, Memory	

IV. CALCULATING THE OPINION SCORE

The main portion of the proposed approach consists of calculating the verb score and adverb score of the tokens present in the three feature arrays extracted from the opinions of the users. Despite the fact that, one can figure out the exact meaning of a specific writings depending on the semantic introduction of the descriptive words, yet including adverb modifiers is very important. This is essentially on the grounds that there are a few adverbs in semantics, (like, *"not"*) which is extremely fundamental to be mulled over as they would totally change the entire meaning of the opinions (i.e.) *"not good"* is a negative view and not a positive view. For example,

User 1: *"The battery life is great"*
User 2: *"The battery life is not great"*

Here by just seeing great, the opinion should not be placed in the positive side as the first user's opinion is positive and the second user's opinion is on the negative side. The next question is if both the views are positive or negative, what is the actual score given for the phrase as every user will have an unique way in expressing their views. For example,

User 1: *"The battery life is great"*
User 2: *"The battery life is extremely great"*

The adverb present in the opinion helps a lot in deciding the level of positivity or negativity and here the User1's opinion is positive, whereas the User2's opinion is strongly positive and a score is provided to find this level as shown in the table 3.

VERB and ADVERB SCORE CHART		
Tokens	**Score**	**Type**
Like, Love,	1	**Verb**
Adore, Enjoy, prefer	0.9	
Excite, Attract, recommend	0.8	
Hate, dislike	-1	
Suck, regret	-0.9	
Suffer, reject, hurt	-.08	
Most, Complete, Best	1	**Adverb**
Too, extremely, very	0.9	
Much, More, pretty	0.8	
Never, Hardly	-1	

Not, less, little	0.8	

Table 3: Common Verb and adverb scores in the opinion

The table 3 showcases most of the frequently employed verbs and adverbs usually in the opinions and their values are ranged from +1 to -1. The Word net is used here to recognize the verbs and adverbs. The token that cannot be identified by the Word net is pruned off as they may not be usable words.

Usually to computerize the user's opinion, plethora of approaches is used to predict the user's sentiments which include Natural Language Processing (NLP) and Machine Learning (ML) algorithms.

The opinion score is calculated after finding the scores of the adjectives and verbs. The adjectives and verbs are combined together to form adjective set and similarly the adverbs and verbs are combined to form verb set. The products of adjective scores, adverb scores and verb scores are initially computed and then the feature based opinion score is calculated by the following equation,

$$S(F_{app}) = \frac{1}{N} \sum_{i=1}^{N} S(Adj) + S(Vb) + (En * S(E))$$

$$S(F_{aff}) = \frac{1}{N} \sum_{i=1}^{N} S(Adj) + S(Vb) + (En * S(E))$$

$$S(F_{ess}) = \frac{1}{N} \sum_{i=1}^{N} S(Adj) + S(Vb) + (En * S(E))$$

Where, S(Fapp) = Score of the feature Appearance, S(Faff) = score of the feature affordability, S(Fess)=Score of the feature essentials, N= number of tokens, S(Adj)= adjective set, S(Vb)= verb set, En=number of emoticons, S(E)= Emoticon score.

V. EXPERIMENTAL RESULT AND ANALYSIS

To explain the score calculation, a sample data is considered and the scores are calculated to find the sentiment involved in the data,

"This mobile is too good :-) and I love the overall design as it is very sleek and light weight. The battery life is great :-) :-). Both the front and the rear camera are the best in the market and the photos are excellent and I enjoy it XD.. But the rate of the phone is very costly :(:(. "

After cleaning the input data, the stop words are removed from it. The basic three features are initially extracted and they are grouped separately to find the score separately. The adjective set and the verb sets are extracted from the separated features data along with the emoticons if present.

Appearance Feature = This mobile is too good :-) and I love the overall design as it is very sleek and light weight.

Adjectives extracted

Too good, very sleek, light weight

Verb extracted

Love

Emoticons extracted

:-)

Affordability feature = But the rate of the phone is very costly :(:(

Essential feature = The battery life is great :-) :-). Both the front and the rear camera are the best in market and the photos are excellent and I enjoy XD

The adjective set and the verb set along with the emoticons are extracted from these features and then the scoring is computed.

(i) $S(F_{app}) = (1 / 6) * ((0.9 \times 1) + (0.9 \times 1) + (0.7 \times 0.5) + (1 \times 0.5))$

$= 0.441$

(ii) $S(F_{aff}) = (1 / 3) * ((0.9 \times -1) + (-0.5) + (2 \times -0.5))$

$= -0.133$

(iii) $S(F_{ess}) = (1 / 7) * ((0.9 \times 1) + (0.9 \times 1) + (0.9 \times 1) + (1 \times 1))$

$= 0.525$

This mobile phone user feedback is classified into three categories and the user has provided a positive opinion about its appearance, positive opinion about its physical nature, but the user has given a negative opinion about its affordability. Overall the opinion may fall on the positive side but the negativity should also be considered and here the rate or the price tag proved to be its negative face. The sample data are shown in the table 4 and its corresponding scores related to appearance, affordability, essentials and the overall rating of the opinion is calculated showcased.

Table 4: Sample data and its score calculations

Sample data	S (App)	S (Aff)	S (Ess)	S (R)
Waste mobile.. internal memory only 8GB can't install more app very very worst X-(	-	-	-0.325	-ve
I am using this phone last 3 month phone is good :D, best battery life.. it's looking and working good but sound quality is not good. Very Cheap phone :-)	0.327	0.212	0.107	+ve
I think it beats all phone in market and it is a best phone to buy at this low price 6 inch screen is very large and good.. I Love this phone. :D :D	0.356	0.211	0.425	+ve

EXCELLENT PHONE :-), I recently purchased the phone in Dubai cheaply whilst on holiday and finally I can say the features are super it is light and classy. My interest in a phone is battery life obviously includes charging speed and clarity of both image and sound. The phone charges to the max in under 30 mins and gives me a lot more :D I LOVE it!. :-)	0.367	0.22 1	0.512	+ve

VI. CONCLUSION

The evolution of online sites selling many products to the consumers offers an exceptional prospect to design and employ plethora of new technologies to search, extract and mine the user's emotions and opinions related to the products that are being sold on the online sites. The work presented in this paper specifies a novel approach for opinion mining on online product data sold. To unearth the sentiments hidden underneath the user fata, the opinion words (a combination of the adjectives along with the verbs and adverbs) are extracted along with the emoticons present in the user sample data. The corpus-based method is employed to discover the adjective set and the dictionary-based approach is employed to discover the verb set. After finding the individual scores of the three important features present, the overall opinion was then found.

REFERENCES

1. Alm, C.O., Roth, D. and Sproat, R. (2005). Emotions from text: machine learning for text based emotion prediction. In Proceedings of the Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP 2005),Vancouver, Canada, p.p. 579-586.
2. A. Kumar. and T. M. Sebastian, "Sentiment Analysis: A Perspective on its Past, Present and Future", International Journal of Intelligent Systems and Applications (IJISA), MECS Publisher, 2012 (Accepted to be published) Bing Liu, Sentiment Analysis, Mining opinions, Sentiments, and Emotions, Book, June (2015).
3. Chunxu Wu, Lingfeng Shen, "A New Method of Using Contextual Information to Infer the Semantic Orientations of Context Dependent Opinions", 2009 International Conference on Artificial Intelligence and Computational Intelligence.
4. Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, "Using Wordnet to Measure Semantic Orientation of Adjectives", Proceedings of 4th International Conference on Language Resources and Evaluation, pp. 1115-1118, Lisbon, portugal, 2004.
5. K.Dave, S. Lawrence, and D.M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews". In Proceedings of the 12 International Conference on World Wide Web (WWW), 2003, pp. 519–528
6. M.R.Saleh, M.T. Martín Valdivia, A. MontejoRáez, and L.A. Urena Lopez, Experiments with SVM to classify opinions in different domains, Expert Syst. Appl. 38, pp. 14799-14804 (2011). WordNet: http://wordnet.princeton.edu/

AUTHORS PROFILE

Mrs.BHUVANESWARI.R Received M.Sc IT Degree in Information Technology from Annamalai university, Chidambaram in the year 2004,received M.Phil Degree in Computer Science from Bharathidasan University Trichy,in the year 2010 persuing (Phd in Computer Sciecne from Bharathidasan University, Trichy)having 10+ year teaching experience and 2 papers published in the international conference. Working as a Asst Professor of Computer Science in Sudharson College of arts and science Pudukkottai, area of Research Data Minining,Big data analysis, Network Security. Eshwari9878@gmail.com. Cell:6369351443,9787872146.

Dr.S.RAVICHANDRAN received his M.Phil degree in computer science from MS University in the year 2004. Ph.D in Computer Science from Bharathidasan University in the year 2014. He is having 20+ years of teaching experience and 7+ years of research experience. He guided more than 50 M.Phil level and guiding 6 candidates at Ph.D level. Right now he is working as HOD and Assit.Prof. in computer science department in H.H.The Rajah's college, Pudukkottai. He published in 20 research paper in international and national level. His research interest includes in the area of Data Mining, Big Data Analytics, Image Processing, AI. rajahsravis@gmail.com.Cell:8838189909.