ANALYZING SENTIMENTS IN PRODUCT REVIEWS

Anshu Jain*

Pragya Shukla**

Hitesh Bandiya***

ABSTRACT

With the vast increase in the utilization of internet in today's era a similar increase has been seen in the use of blog's,reviews etc. The person who actually uses these reviews or blog's is mostly a consumer or a manufacturer. As most of the customers of the world are buying & selling product on line so it becomes company's responsibility to make their product updated. In the current scenario companies are taking product reviews from the customers and on the basis of product reviews they are able to know in which they are lacking or strong this can be accomplished with the help of sentiment analysis. As Opinions Play important role in the process of knowledge discovery or information retrieval and can be considered as a sub discipline of Data Mining .In this research effort we demonstrate a technique which is based on rules where product reviews are extracted from review containing sites and analyze them so that a person may know whether a particular product review is positive or negative or neutral. The system will utilize a existing knowledge base for calculate positive and negative scores and on the basis of that decide whether a product is recommended or not.

Keywords: *Opinion Mining,Sentiment Analysis,Sentiment Orientation,Opinion Analysis*

*Institute of Engineering & Technology, Indore, India.

**Assistant Professor, Institute of Engineering & Technology, Indore, India.

***Institute of Engineering & Technology, Indore, India.

I. INTRODUCTION

Opinion are the views that are expressed by an individual on some topic/issue according to his/her own perspective. As these views have been used by several application domains such as business and organization,individual etc,we can say it has become very important to find out efficient ways for extracting opinions. So Opinion mining is the process of studying people's opinions or emotions towards entities, events and their features. In the past few years, it has attracted a great deal of attentions from both academia and industry due to many challenging research problems and a wide range of applications. Opinion mining deals with the sentimental words which clearly demonstrates ideas about the targeted object at that particular time. Opinion mining is a type of natural language processing for gathering the information from public about a particular product. Whenever we decide to buy something new we always look forward for someone's advice for same Opinion mining has wide range of application due to its increasing popularity among the public. for example, it can help you judge that which product is good and which is bad while both are having the same price,market analysts may use the reviews posted in various blogs for taking major decisions regarding new product launch. The another name for opinion mining may be sentiment analysis and subjective analysis. The. objective of opinion analysis is to identify emotions from text and determine their polarities. On the basis of determined polarities we can conclude whether the text document represents positive ,negative or neutral opinion . The discovered opinions are useful to many practical applications such as opinions in product quality reviews are helpful to potential customers. Meanwhile,opinion mining technique relates indirectly to promote many natural language processing techniques.

A lot of research has currently been done over this emerging technology but our work is focused on two things,firstly we extract the feature words in a given review and secondly the summary will be produced by assigning scores to the feature extracted words in first step,in this step we also display the opinion words that are responsible for positive or negative biasing of reviews Remaining part of the paper has been organized in the following manner:- part II describes the involve resources for conducting this research .part III gives an overview of proposed Technique for opinion summarization,part IV describes the system design and part V concludes this paper.

II. INVOLVED RESOURCES

In this section we describe the necessary resources involved in summarizing the opinions present in product reviews. As our work is concentrated on assigning polarity score to various

feature words which are extracted we need to use some existing lexical resource which must be effective and can smartly extract positive and negative score for extracted words. Here comes the utility of SentiWordnet given By A. Esuli and F. Sebastiani,[10],It is described as below:-

(a) SentiWordNet:-it is a lexical resource of sentiment information for terms in the English language designed to assist in opinion mining tasks,where each term is associated with numerical scores for positive and negative sentiment information. It provides a readily available database of term sentiment information for the English language. Each term in SentiWordNet is associated with numerical scores for positive and negative sentiment information.SentiWordNet can be a replacement to the process of manually deriving lists of terms containing sentiment information for opinion mining tasks.

Another resource which is utilized in this research is POS Tagger for part-of-speech tagging purposes which is   freely provided by Stanford university for research purposes. This Described as below:-

(b) Stanford POS Tagger:-The Process of assigning different parts of speech tags such as noun,adjective and adverbs to a given text is known as Part-Of-Speech tagging.

A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.   This software is a Java implementation of the log-linear part-of speech taggers. The system requires Java 1.5+ to be installed.

The English taggers use the Penn Tree bank tag set. Here are some tags of the Penn Tree bank English POS tag set:-

Sno.	Tag	Description
1	NNP	noun, proper, singular
2	NNPS	noun, common, plural
3	RB	adverb
4	JJ	adjective or numeral, ordinal
5	JJR	adjective, superlative
6	NN	noun, common, singular

7	RBR	adverb, comparative
8	VB	verb, base form
9	VBD	verb, present participle or gerund

Table 1 Tag set Description

III. PROPOSED TECHNIQUE

In this section we describe our proposed algorithm for the effective extraction of useful information from product reviews. This algorithm works on the basis of   rule based methodology using lexicon approaches as described in the above sections. For better accomplishment of the steps we need to take some assumptions in advance .so we consider that the product reviews from various review relevant sites have already been extracted with the help of appropriate procedures and are already stored inside the system. Therefore we skip the task of extracting the reviews and storing them in our system which requires some sort of crawlers to get it done. Then after we take a set of product reviews as input and consider a polarity score along with the relevant feature words as expected outcome. Following are the number of steps required to generate the expected outcome:-

Proposed Algorithm

INPUT:Set of product reviews

OUTPUT:Extracts opinion words with positive and negative polarity

BEGIN

1) Collect the reviews Rn for various product Pn

2) Each review Ri contains only one sentence in one line

3)   (a) For each Review Ri in

         For each Sentence Si in R

         For each word Wi in Sentence Si

        (b) Check whether the word is in stop word list eg   A,and,the,for,at etc

        If found then replace it with white space

        (c) Assign P_no and POS tag to each word.

        (d) Extract the patterns which matches predefined   rules.

(e) Assign positive and negative score to each word in extracted pattern. Using

Function assign_score()

If the pos score>neg score then

Fea_Word score=pos score

Else if pos score<neg score then

Fea_Word score=neg score

(4) Repeat step 3 (a) and check whether the word matches any word from Conditional dictionary.

Eg negative words such as not,never,neither etc

If found then recalculate its score

Fea_Word score=Fea_Word score+Cond_weight

5) Sentence score will be calculated using the

$$i=1$$

Sent_Score=∑Fea_Wordscore(1)+Fea_Wordscore(2)+..   …... +Fea_Word score(n)

              n   n

(6) Finally overall review score will be

$$i=1$$

  Review_score=∑ Sent_Score(n)

              n   n

7) Calculate Accuracy using recall ,precision and F-measure

precision[Positive] =True Positive / True Positives + False Positives

recall[Positive] =True Positives/True Positives + False Negatives

F-measure- it is the harmonic mean of precision and recall. It is computed using the formula defined below.

 F − measure = 2 *(precision * recall /precision + recall)

As mentioned in first step of algorithm we retrieve previously extracted reviews and provide them as input to our system. In the next step for accomplishing sentence level sentiment classification we   process each sentence in a individual review and then each word in a individual sentence for further processing. Third step comprises of loop control structure for review,sentence and word respectively. Starting at word level we check whether the any word matches the word in stop word list and if does then replace it with whitespace .The reason behind removing these words is that they contain very less probability of holding any kind of opinion information. After this within the same step we assign position to each word and perform parts of speech tagging of remaining words with the help of existing resources as mentioned in section 2.

In step 3 (d) we extract useful patterns which are the collection of most probable words those are considered to be good polarity indicators. So the useful patterns which we used for our methodology are as below:-

Pattern	First word	Second Word	Third word
P1	Adjective(good)	Noun(Image)	Noun(Quality)
P2	Verb (Recommended)	Noun(camera)	-----
P3	Adverb(very)	Adjective(short)	Noun(life)
P4	Adverb(not)	Adverb(very)	Adjective (good)

Table 2 Description of predefined Pattern rules

After sucessful extraction of predefined rules the resulting words will be assigned to the positive and negative score with help of knowledge dictionary and Function assign_score() will decide whether a word is assigned with positive or negative score.Each word in the pattern will get the positive or negative score.Nextly step 4 again performs the matching each word in input review with the conditional dictionary if wor found then recalculation of polarity score is done .Conditional dictionary contains contextual word and due their presence polarity will get affected so we perform dual checking to detect these words.Rest of the steps 5,6,7 Contains the formula for calculating scores for a word,sentence and review respectively.

IV. SYSTEM DESIGN

Figure 1 represents our proposed system design,where task of opinion mining is divided in to two parts, one is data preprocessing and other is data main processing. In data preprocessing we firstly remove stop words from input text so as to be more focused on required data and then parts of speech tagging for further processing of remaining text. In data main processing

we firstly extract the predefined phrases and then polarity score is assigned to them .Nextly we evaluate our technique By calculating recall and precision over obtained results.
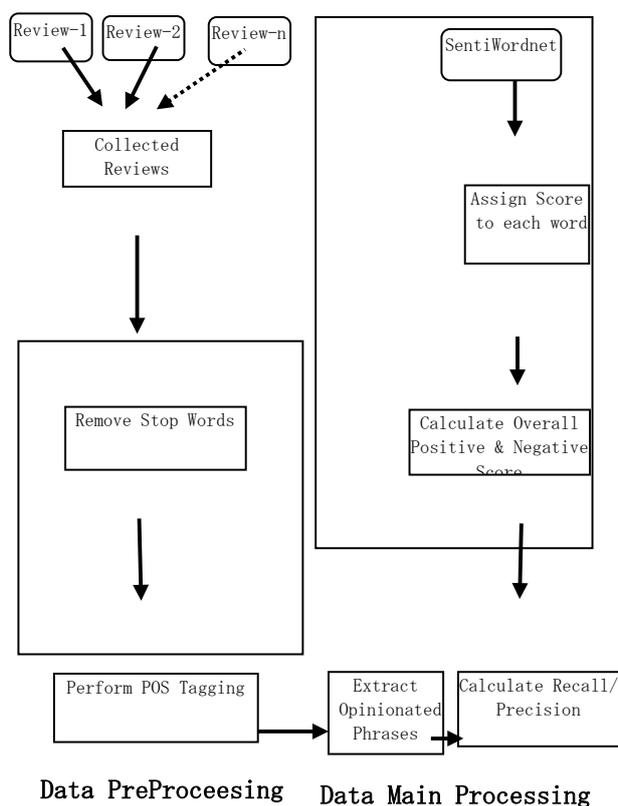

Figure 1 Proposed System design

V. CONCLUSION

Opinion Mining   is a emerging technology applying to but yet need get improvement to achieve more convenient and it is emerging and rapidly growing field .however the previous research work mainly focused on training dataset for analyzing the sentiments present in user generated content, they succeed in providing good level of accuracy but due to lack of availability of annotated corpus for same purposes,we emphasis on the utility of lexical resources ,In this paper we showed the usability of these sources over supervised learning . Opinion Mining has become Important for all types of fields in extracting likes and dislikes and intensity of likes and dislikes. So the research conducted here is hopefully of significant use to help the researchers in knowing about the past and recent trends in the same area.

REFERENCES

[1] B. Liu. Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010.

[2] B. Liu,Sentiment Analysis: A Multi-Faceted Problem IEEE Intelligent Systems,2010

[3] B.Liu,"Opinion Mining",invited contribution to Encyclopedia of Database Systems,2008

[4]Hu and Bing Liu, "Mining Opinion Features in Customer Reviews, Mining", American Association for Artificial Intelligence 2004.

[5] Bo Pang and Lillian Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval Vol. 2, Nos. 1–2 (2008) 1–135 2008.

[6]Minqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews", KDD'04, August 22–25, 2004, Seattle, Washington, USA, 2004.

[7]Ruifeng Xu1,2, Kam-Fai Wong2, Qin Lu1, Yunqing Xia3, Wenjie Li1"Learning Knowledge from Relevant Webpage for Opinion Analysis,2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology,2008

[8]A. Esuli,"Opinion Mining",Language and Intelligence Reading Group,June 14, 2006, Pisa, Italy.

[9]Martin, J. R. and White, P. R. R. (2005).The Language of Evaluation: Appraisal in English.

[10]A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in Proceedings of Language Resources and Evaluation (LREC), 2006.

[11]A. Esuli and F. Sebastiani, "PageRanking WordNet ,synsets: An application to opinion mining," in Proceedings Q. L. Miao, Q. D. Li, and R.W. Dai,A sentiment mining and retrievalsystem. Expert Systems with Applications: An International Journal,2009,36 (3):7192-7198

[12]H. Yu, and V. Hatzivassiloglou, Towards answering opinion questions:Separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. Morristown, NJ, USA:Association for Computational Linguistics, 2003, pp. 129 – 136.

[13]Xiaojun Li,Lin Dai,Hanxiao Shi,"Opinion Mining of Camera Reviews Based on Semantic Role Labeling",2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)

[14]Turney, P. (2002).Thumbs up or thumbs down? Semantic orientation applied to unsupervisedclassification of reviews.In Proceedings of ACL-02, 40th Annual Meeting of the Association forComputational Linguistics, pages 417–424. Association for Computational Linguistics.

[15]Turney, P. D. and Littman, M. L. (2003).Measuring praise and criticism: Inference of semantic orientation fromassociation.ACM Transactions on Information Systems, 21(4):315–346

[16].Liu,Sentiment Analysis: A Multi-Faceted Problem IEEE Intelligent Systems,2010