Enhanced Sentiment Analysis of Informal Textual Communication in Social Media By Considering Objective Words And Intensifiers

Jasmine Bhaskar
Department of Computer Science and Engineering
Amrita Create
Amrita Vishwa Vidyapeetham
Amritapuri, India
jasmine@am.amrita,edu

Sruthi K
Department of Computer Science and Engineering
Amrita Create
Amrita Vishwa Vidyapeetham
Amritapuri, India
sruthikallath@gmail.com

Prema Nedungadi
Department of Computer Science and Engineering
Amrita Create
Amrita Vishwa Vidyapeetham
Amritapuri, India
prema@am.amrita,edu

Abstract*— Sentiment analysis is a valuable knowledge resource to understand collective sentiments from the web and helps make better informed decisions. Sentiments may be positive, negative or objective and the method of assigning sentiment weights to terms and sentences are important factors in determining the accuracy of the sentiment classification. We use standard methods such as Natural Language Processing, Support Vector Machines and SentiWordNet lexical resource. Our work aims at improving the sentiment classification by modifying the sentiment values returned by SentiWordNet for intensifiers based on the context to the semantic of the words related to the intensifier. We also reassign some of the objective words to either positive or negative sentiment. We test our sentiment classification method with product reviews of digital cameras gathered from Amazon and ebay and shows that our method improves the prediction accuracy.

Keywords— ***Sentiment analysis, opinion mining, sentiment polarity, subjective words, objective words, intensifier***

I. INTRODUCTION

Sentiment analysis has become a new knowledge resource after the advent of the Internet and World Wide Web. It aims to automatically predict the sentiment polarity of user's opinions on the web. Opinions play an important role in understanding the collective sentiments and help to make better decisions. Opinions may be positive, negative or neutral. Positive opinions encourage the prospective customer to take positive decision; negative opinion usually results in negative decision. Sentiment analysis of textual communication extracts the subjective information in the text.

The main task in sentiment classification is to determine the polarity of the comments as positive, negative or objective. It can be done at different levels such as word/phrase levels, sentence level and document level. Sentiment analysis is one of the most challenging areas in NLP because people express opinion in subtle and complex ways, involving the use of slang, ambiguity, sarcasm, irony and idiom.

Most of the research in the field of sentiment classification focuses on polarity classification of review documents. One of the main tasks in sentiment classification is the discovery of sentiment words. This task can be done more easily by applying sentiment lexicons such as General lnquirer, SentiwordNet, WordNet Affect, SenticNet. Once the sentiment words are identified, the score of the sentences can be calculated.

Majority of the existing models of sentiment classification ignores objective words and intensifiers during score calculation. An intensifier enhances the emotional context of the word it is tagged by increasing the degree of the adjective or adverb. These adjectives or adverbs may be positive or negative sentiment words. The word 'totally ', 'pretty ', 'incredibly ', 'very ', 'really 'etc are example for intensifiers. Most of the intensifiers obtain positive score from the SentiWordNet, a publicly available lexical resource used for sentiment classification [1]. So during score calculation, negative sentences will not get the real effect of intensifiers

This paper intends to improve the sentiment classification by modifying the sentiment values returned by SentiWordNet for intensifiers based on the context to the semantic of the words related to the intensifier. We also consider objective words and reassign them to either positive or negative sentiment depending on a threshold value.

The rest of the paper is organized as follows: Background and Related works is given in Section 2, Solution approach in Section 3, Experiment Results in Section 4 and finally the concluding remarks in Section-5.

II. BACKGROUND AND RELATED WORK

During the initial stages of sentiment analysis, researchers were more focused on subjectivity detection. Over a decade or two, due to the rapid growth of World Wide Web and internet, the web became the main source for gathering information. Hence the researchers gradually shifted their focus from subjectivity analysis to sentiment analysis of online reviews [15]. Now a day's people express their opinions on social media which consist of product review sites, social networks, blogs, or forum (such as Amazon, Face book, Twitter, Flicker, LinkedIn, etc.). Information from these sources is really helpful for the customers to take purchase decisions.

[978-1-4799-4040-0/14/$31.00 ©2014 IEE]

The major tasks in sentiment classification are the extraction of sentiment words and calculation of sentence scores. The task of extracting sentiment words will be tedious and exhaustive if do manually. This task can be performed more easily by applying sentiment lexicon. SentiwordNet is one of the most powerful lexical resources used for sentiment classification [1].

Listed below is some works done in sentiment analysis using SentiWordNet.

A. Sentiment classification using SentiwordNet

Ohana and Tierney[7] proposed a technique for sentiment classification by using features built from the SentiWordNet database of term polarity scores. Their approach consists of counting positive and negative term scores to determine the sentiment orientation. They also presented the improved version of this by building a data set of most important features using SentiWordNet and a machine learning approach for classification. A negation detection algorithm is used to adjust the SentiWordNet scores accordingly for the negated terms.

Ghosh and kar[15] described a simple approach to perform sentiment classification based on an unsupervised linguistic approach. This works use SentiWordNet for classifying the online reviews. They have done feature selection based on sentence tagging and then identified the subjective sentence based on the opinion words. Then the positive and negative score of each sentence is calculated. The results of this work show that SentiWordNet could be used as an important resource for discovering sentiment words and classification tasks.

Denecke[6] proposed a technique for determining the polarity of text within a multilingual framework. In this work, at first the documents other than English are converted into English by using standard translation software. Translated document belongs to one of the two classes 'positive' or 'negative' according to its sentiment score. The score of the sentiment bearing words are calculated using SentiWordNet. For classifying the documents according to their sentiment (positive, negative), three different approaches were implemented and evaluated. The results of these three methods were then compared. Best result was obtained by using the machine learning technique.

Hung and Lin[14], introduced an approach for sentiment classification by considering objective words. They assign a new value to objective words. Basic concept is positive or negative sentence has sentimental impact on its affiliated objective words. If an objective word appears in more positive sentences than a negative sentence, then it has a positive polarity and vice versa. An objective word is not changed when it appears in both positive and negative sentence equally. Reassigned value of the objective word could be meaningful ie, it can have if either positive or negative sentiment value which is greater than a threshold value 0.5.

B *SentiWordNet*

SentiwordNet [1] is a publicly available lexical resource explicitly used for sentiment classification and opinion-mining. SentiWordNet is based on WordNet (version 3.0). SentiWordNet assigns sentiment scores to each synset of WordNet: positive, negative or objective

WordNet terms are mainly categorized into four: Adverb, Adjective, Noun and Verb. Terms are linked by meaning and grouped into synset (synonymous sets).Each synset has three numerical scores: pos(s), neg(s), obj(s) ranging in the interval [0,1]. Sum of score pos(s), neg(s) and obj(s) is equal to one. So obj(s) can be calculated as obj(s) = 1 - (pos(s)+neg(s)).

Figure1 shows the graphical model designed by Esuli and Sebastiani [1] to display the scores of a synset in SentiWordNet.

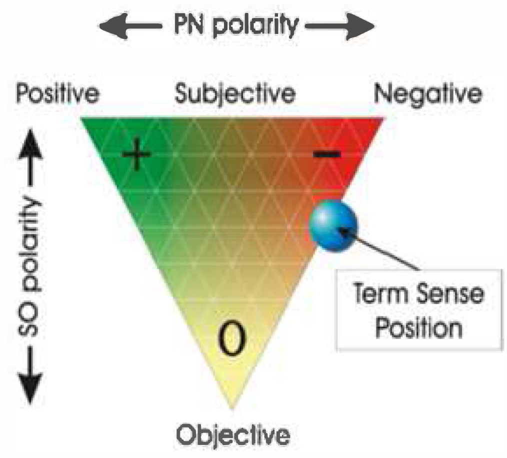

Fig. 1. Graphical Representation for SentiWordNet [1]

II. SOLUTION APPROACH

In this work, we propose an enhanced technique for sentiment classification of online reviews by considering the objective words [5] and intensifiers. The proposed work consists of three major modules as shown in Figure2.

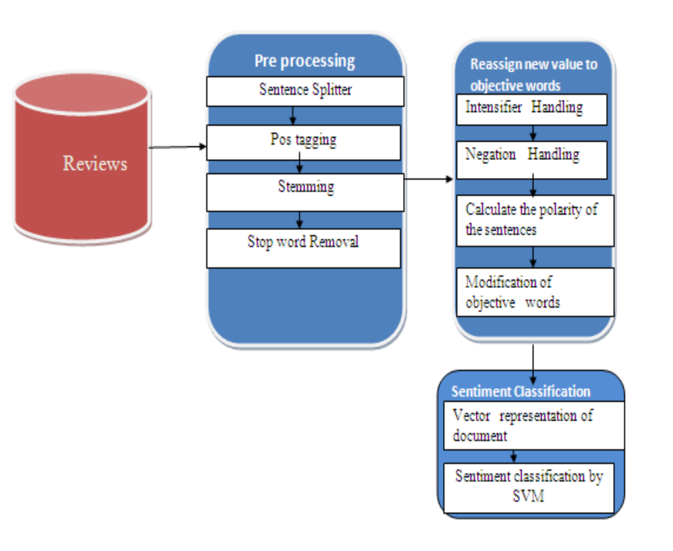

Fig. 2: Frame work of proposed model

A. Review Document Preprocessing

Preprocessing follows the same step as traditional text mining which consists of sentence splitter, POS tagging, stemming and stop word removal. Review document consists of many sentences and each sentence expresses specific sentiment. So sentence is considered as a basic unit here. Review documents are first split into several sentences based on punctuation such as semicolons, question mark, exclamation point or period.

Review document contains so many spelling mistakes. We used enchant spellcheckers library to correct them.

SentiWordNet provides four POS classes: Adjective, Adverb, Verb and Noun. Same word with different part of speech tag might have different sentiment value. For example Word 'good' appears in three different parts in a sentence may have different values according to its part of speech tag. So proper part of speech tag should be applied on each word in the sentences.

As all words are stored in SentiWordNet in its base form stemming are required for each word in the sentence. Stemming is the process of conversion of a word to its base form.

Stop words are the word that doesn't carry much meaning such as determiners and prepositions. Removal of stop words is the last step in preprocessing. In this work NLTK [16] toolkit in python is used for preprocessing.

B. Reassign new value to objective words

This module consists of four steps.

- Negation Handling
- Intensifier handling
- Calculate the polarity of the sentence
- Modification of objective words

Negation Handling First step in this module is a negation processing. Suppose a sentence contains negative modifiers like 'not ', 'never', 'no' which change the meaning of that sentence. For example consider the sentence: 'The picture quality of this camera is not bad'. This sentence contains a negative modifier 'not' which changes the negative polarity of the sentence to positive polarity. If we calculate the score of the sentence without handling negation, we get a high negative polarity as 'not' and 'bad'.

Intensifier Handling People usually use intensifiers in reviews to express their emotion deeply. Presence of the words like 'very ', 'really 'and 'extremely ' in negative and positive sentences make the adjective and adverb stronger. But this effect is not considered during the score calculation in existing method.

For example, the score of the sentence, 'Camera is bad 'when calculated using the existing method gives -0.380527. When intensifier 'very' is considered, the new score became 0.1194 as 'very' is considered as a positive word. Score of the sentence 'Camera is very bad' should be more negative as compared to sentence 'Camera is bad'. But we obtained a positive score from the existing method.

Table1 shows how to handle intensifiers in positive and negative sentences.

TABLE I: INTENSIFIER HANDLING IN POSITIVE AND NEGATIVE SENTENCES

Previous Word	Next Word	Score
Intensifier	Adjective[Negative]	High Negative
Intensifier	Adjective[Positive]	High Positive
Intensifier	Adverb[Negative]	High Negative
Intensifier	Adverb[Positive]	High Positive

The polarity of the sentence can be obtained by equation (1)

$$SentenceScore = \sum_{i=0}^{n} score(i) \qquad (1)$$

Score(i) is the positive and negative score of the words and

n is the number of words in the sentence. If Sentence

Score is greater than 0, then we can say that the sentence is positive otherwise sentence is negative.

Consider the following examples.

Sentence1: Camera is bad (p=0,o=0.62 n=0.38) and

Sentence2: Camera is very (p=0.5 o=0,n=0) bad(p=0,o=0.62 n=0.38) .

Here' is' is a stop word which has no score in SentiwordNet .Camera is not a sentimental word.

Sentence1 contains only one negative sentiment word 'bad '. Hence Sentence1 is negative Sentence2 contains one positive word and one negative word. So score of the sentence2 obtained by using equation1 is positive. Actually Sentence2 has more negative polarity as compared to Sentence1. But this effect is not reflected as most of the intensifiers have positive score in the SentiWordNet. . Algorithm1 is the proposed algorithm for handling intensifiers.

Algorithm 1 Intensifier Handling

```
If (Score (Nextword)>0) then
  (Take the actual score of the intensifier)
Else (Score (Nextword)>0) then
   If previous word is Negative Modifier then
     (Take the actual score of the intensifier)
   Else
     (Negate the score of the Intensifier)
  End if
End if
```

For example score of the sentence, `camera is bad' calculated using proposed method gives -0.380527. When intensifier very is considered, score became -0.880527.Here the sentence 'camera is very bad' obtained high negative score as compared to sentence1.

After negation processing and intensifier handling is done, next task is to calculate the polarity of sentences. Depending upon the sore of the sentence, we can say whether the sentence is positive or negative.

Last step in this module is reassigning new value to objective words. Almost 90% of the words in SentiWordNet are objective words. Most of the existing models simply ignore the objective words [11]. But the objective words have sentiment influence on the associated words. So a new value is reassigned to the objective words. If an objective word appear in a more positive sentence as compared to negative sentence then it has positive polarity otherwise it has a negative polarity. An objective word is not reassigned if it appears in both positive and negative sentences equally.

Let's assume an objective word appear in positive, negative, objective in eight, one, one times respectively. Then a new value 0.8 is assigned to the objective word

In Algorithm2, $Posw_i$ indicate value in positive orientation $Negw_i$ indicates a value in negative orientation and $Objw_i$ indicate value in objective orientation. ps_i indicate frequency

of wordi in positive sentence. ns_i indicate frequency of $word_i$ I n negative sentence. fr_i indicate term frequency of $word_i$ in the data set.

Algorithm 2 Reassigning new value to objective words

if(Objective word occurs only in positive sentence) **then**
$Posw_i = ps_i/fr_i$; $Negw_i=0$; $Objw_i=1-Posw_i$
else if(Objective word occurs only in negative sentence)**then**
$Negw_i = ns_i/fr_i$; $Posw_i = 0$; $Objw_i=1-Negw_i$
else if(Occurrence of positive sentence<Occurrence negativesentence) **then**
if(nsi-psi>theroshold) **then**
$Negw_i = ns_i/fr_i$; $Posw_i = 0$; $Objw_i=1-Negw_i$
end if
else if(Occurrence of positive sentence>Occurrence negative sentence) **then**
if(ps_i-ns_i>theroshold) **then**
$Posw_i=ps_i/fr_i$; $Negw_i=0$; $Objw_i=1-Posw_i$
end if
else
$Posw_i=0$; $Negw_i=0$; $Objw_i=1$
end if

A sentence or a word containing a smaller positive or negative sentiment value won't have a great effect on its associated objective words. Our modification is meaningful only if $Posw_i$ or $Negw_i$ is greater than threshold. Here we assign a threshold as 0.5.

C Sentiment Classification

This module consists of two steps. Vector representation of the document and sentiment classification using SVM. SVM has high accuracy as compared to other machine learning approaches [3].

At first, represent the document as a vector

$D_i = [W_1, W2, W3...W_n]$

Where W_i is the weight of the term i with respect to the document. If a term W_1 occurs in sentence $_i$, then W_i will have a non zero value, otherwise it is zero. Document matrix has each row_i correspond to the feature vector of a particular sentence, and each column of this matrix refers to unique term (feature) in the document.

Weight of the positive word is calculated as.

$$W_i=TF_i * Posw_i$$

Weight of the negative word is calculated as.

$$W_i=TF_i * Negw_i * -1$$

Weight of the objective word assigned as $W_i=0$

Where W_i is the weight of $Word_i$ and TF_i is term frequency of the word in the dataset.

IV EXPERIMENT RESULTS

In this section, we demonstrate the performance of our proposed method by comparing it with the existing method. For our experiments we have used the data set taken **from Amazon.com** and **ebay.com** for the products reviews of digital camera from which we have taken **24000** sentences for training and **20000** sentences for testing.

The performance of the classifier can be measured in terms of the four possible outcomes: True positive (TP), true negative (TN), false positive (FP), and false negative (FN). True positive\negative means that a sentence is classified to a positive\negative class when this sentence really belongs to the positive\negative class respectively. Both true positive and true negative are correct classifications. False positive\negative means that a sentence is incorrectly classified to a negative\positive class when this sentence belongs to a positive\negative class.

Accuracy of existing and proposed method is calculated by using the equation (2) given below.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

TABLE II: SENTENCE TAKEN FOR TRAINING AND TESTING

Training	Testing

24000	20000

Table III shows the performance of the traditional, existing and proposed methods and lists the accuracy, the true positives, true negatives, false positives and false negatives of sentences.

From the above result we can see that the prediction accuracy of the proposed method which is much better than the traditional and existing methods. Reassigning objective word as positive or negative, improved the classification accuracy by reducing both positive and negative misclassification. Though the existing method out performs the traditional method its accuracy is less compared to the proposed method. This is because miss-classification is less in the proposed method related to the negative sentences as compared to the existing method. This improvement is due to the proper handling of intensifiers.

TABLE III: PERFORMANCE OF THE TRADITIONAL, EXISTING AND PROPOSED METHODS

	Traditional Method Non revised SentiWordNet	Existing Method Reassigning objective words	Proposed Method Reassigning objective words and intensifier handling
Classification Accuracy	71.095	74.195	76.02
True Positives	7178	7493	7493
True Negatives	7203	7353	7488
False Positives	2797	2647	2512
False Negatives	2822	2507	2507

CONCLUSION

In this paper we have proposed a new method to improve the sentiment classification of product reviews by considering the objective words and intensifiers. Here we have used SentiWordNet, a publically available lexical resource to extract the sentiment word from the online reviews.

After performing pre processing, negation handling, intensifiers handling and modification of objective words, polarity of the sentences is calculated. In this work we have used SVM for sentiment classification. Our proposed model can handle classification in an efficient manner than traditional methods. Our experiment results show that proposed approach achieved better accuracy than existing methods. This is proved by the experiments performed on the product reviews of digital camera. In this work word sense disambiguation are not considered. Word sense disambiguation and identification of the product feature about which the sentiment is expressed can be done as a future work.

ACKNOWLEDGMENT

This work derives direction and inspiration from the Chancellor of Amrita University, Sri Mata Amritanandamayi Devi. We thank Dr. Ramachandra Kaimal, head of Computer Science Department, Amrita University for his valuable feedback.

REFERENCES

[1] A. Esuli and F. Sebastiani, SentiWordNet:A Publicly Available Lexical Resource for Opinion Mining, *Proc. 5th Int Conf. Language Resources and Evalua-tion, European Language Resources Association* (ELRA)2006, pp. 417422.

[2] C. Hung, C.-F. Tsai, and H. Huang,Extracting Word-of-Mouth Sentiments via SentiWordNet for Document Quality Classification ,*Recent Patents on Computer Science* vol. 5, no. 2, 2012, pp. 145152.

[3] Bo Pang and Lillian Lee, Shivakumar Vaithyanathan , Thumbs up?Sentiment Classification using Machine Learning Techniques *Proceedings of EMNLP*. 2002.

[4] B. Liu, Sentiment Analysis and Subjectivity, Handbook of Natural Language Processing *Handbook of Natural Language Processing, 2nd ed.,N. Indurkhya and F.J.Damerau, eds*, Chapman & Hall CRC Press, 2010

[5] H. Saggion and A. Funk, Interpreting SentiWordNet for Opinion Classification, *Proc. Intl Conf. Language Resources and Evaluation*, ELRA, 2010, pp.11291133.8

[6] Kerstin Denecke. Using SentiWordNet for multilingual sentiment analysis *In ICDE Workshops [DBL08*] pages 507-512.

[7] B. Ohana and B. Tierney, Sentiment Classification of Reviews Using SentiWordNet, *Proc. 9th IT&T Conf., Dublin Inst. technology*, 2009.

[8] Z. Zhang, X. Li, and Y. Chen, Deci-phering Word-of-Mouth in Social Media: Text-Based Metrics of Consumer Reviews, *ACM Trans. Management Information Systems*, 2009.

[9] H. Tang, S. Tan, and X. Cheng, "A Survey on Sentiment Detection of Reviews," *Expert Systems with Applications*, PergamonPress, 2009, pp. 10760–10773.

[10] S. Baccianella, A. Esuli, and F. Sebastiani, SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, *Proc. Intl Conf. Language Resources and Evaluation.* 2010, pp. 22002204.

[12] E. Cambria, C. Havasi, and A. Hussain, SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis, *proc. 25th Intl Florida Artificial Intelligence Re- searches Society Conf* AAAI, 2012, pp. 202207.

[13] B. Heerschop et al., Polarity Analysis of Texts Using Discourse Structure,. *Proc. 20th ACM Intl Conf. Information and Knowl-edge,Management*, ACM, 2011, pp. 10611070

[14] Chihli Hung and Hao-Kai Lin, Chung Yuan Christian University,Using Objective Words in SentiWordNet to Improve Word-of -Mouth Sentiment Classifications.*IEEE Intelligent Sys,tems* vol. 28, no. 2, pp. 47-54, March-April 2013

[15] Monalisa Ghosh, Animesh Kar, Unsupervised Linguistic Approach for Sentiment Classification from Online Reviews Using SentiWordNet 3.0, *International Journal of Computational Linguistics*, vol.20, 1994, pp. 233-287.

[16] Bird, Steven; Ewan Klein; Edward Loper (2009), Natural Language Processing with Python, *O'Reilly Media Inc*, ISBN 0-596-51649-5.