An Improved Sentiment Classification for Objective Word

Sangita N. Patel[1] Ms. Jayna Shah[2]
[1]M.E. Student [2]Assistant Professor
[1,2]Department of Computer Engineering
[1,2]Sardar Vallabhbhai Institute of Technology, Vasad-388306, Gujarat-India

Abstract— Sentiment classification is an ongoing field and interesting area of research because of its application in various fields. Customer sentiments play a very important role in daily life. Currently, Sentiment classification focused on subjective statements and ignores objective statements which also carry sentiment. During the sentiment classification, problem is faced due to the ambiguous sense (meaning) of words and negation words. In word sense disambiguation method semantic scores calculated from SentiWordNet of WordNet glosses terms. The correct sense of the word is extracted and determined similarity in WordNet glosses terms. SentiWordNet extract first sense of word which used in general sense. This work aims at improving the sentiment classification by modifying the sentiment values returned by SentiWordNet and compare classification accuracy of support vector machine and naïve bays.

Key words: Sentiment classification, Word sense disambiguation, SentiWordNet, WordNet

I. INTRODUCTION

Text documents from various sources within the publication of this research for the discovery of the overall opinion shows an initial proposal. Examples of these sources, news articles, blog posts, social media, movie review, game reviews and product reviews. Furthermore, opinion mining also called sentiment analysis , recent interest has gained more than a decade, which is a relatively new field of study. This kind of positive and negative meaning to categorize groups wrote documents. Can also be neutral to the mentioned but as having no opinion or cancel each other that the same amount of positive and negative thoughts that can be see there. Computers are getting faster on the Internet, an increasing amount of data is becoming more available to everyone. Very large amounts of information are available in on-line documents.

Sentiment analysis is a process for tracking the mood of the public about a certain product, for example, by building a system to examine the conversations happening around it. As part of the effort to better organize this information for users, researchers have been actively investigating the problem of automatic text categorization [1]. A demand to extract and available to obtain useful information from the data for data mining that enables individuals and companies have risen. It is also the case with sentiment analysis. People's opinion about companies, movies, games, education, products and people is a way to identify the needs. For example, companies began to brand its new product was received in the media how the image. It also search for relevant products to the order of the options available to them by public opinion to give a holistic approach can be used to help people. Gardening is talking about, when the plants needed as human beings do not enjoy similar weather conditions, as the sentence has meaning can be completely opposite. Opinions can be expressed in different ways. The following are examples of statements of opinion.

- I read this book.
- The book is good.
- I like to read this book.
- The book is very good.
- The camera's quality is very good.

In sentiment analysis, researchers have focus mainly on two problems in detecting whether subjective or objective text, and subjective text is further classify as positive or negative. Ask sense orientation, calculation and supervised and unsupervised classification based on machine learning techniques: Techniques used two main approaches. Opinions varied sources, such as can be gathered from personal conversations, newspapers, television, Internet, etc., the Web has become the richest source of feedback collection. Before World Wide Web (WWW), people collected opinion manually.

Sentiment analysis has become a new knowledge resource after the advent of the Internet and World Wide Web. It aims 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. Word Sense Disambiguation process is a process to define the sense/meaning of an ambiguity word. The sense of a word in a text depends on the context in which it is used. The context is determined by the other words in the neighborhood in the sentence. To give a hint how all this works, consider two examples of the distinct senses that exist for the (written) word "bass":

- a type of fish
- tones of low frequency and the sentences:
- I went fishing for some sea bass.
- The bass line of the song is too weak.

To a human, it is obvious that the first sentence is using the word "bass (fish)", as in the former sense above and in the second sentence, the word "bass (instrument)" is being used as in the latter sense below[2].

II. LITERATURE SURVEY

This section describes literature review or the studies which give an idea that for our research done in direction of sentiment classification.

Yan Dang, Yulei Zhang proposed lexicon enhanced method for sentiment classification combines machine learning and semantic-orientation approaches into one framework that significantly improves sentiment classification performance. We also found that conducting feature selection can further improve the performance, especially for large data sets. They compared Naïve Bayes, Maximum Entropy, and SVM and achieved the highest classification accuracy (82.9 percent) using SVM[4].The semantic-orientation approach, on the other hand, performs classification based on positive and negative sentiment words and phrases contained in each evaluation text and mining the data requires no prior training.[4]

Chihli Hung,Hao-Kai Lin proposed approach for mine sentiments of opinions from word-of-mouth (WOM) to improve the performance of word-of-mouth Sentiment classification by re-evaluates objective sentiment words in the SentiWordNet sentiment lexicon with the help of SVM classifier.[5] WordNet is a public sentiment lexicon that's used to extract sentiments of WOM for sentiment classification. However,most existing sentiment mining models ignore objective words, which comprise more than 90 percent of the words in SentiWordNet. These objective words are often considered useless. Research reevaluates objective words in SentiWordNet by assessing the sentimental relevance of objective words and their associated sentiment sentences. In this paper two sampling strategies and integrate them with the support vector machines (SVMs) for sentiment classification.[5]

As an example, we'll use two sentences wherein each word contains three sentiment values in brackets—that is, Positive, objective, and negative—while looking up SentiWordNet as follows:

- Sentence 1: I (p:0, o:1, n:0) will (p:0, o:1, n:0) read (p:0, o:1,n:0) this (n/a) book (p:0, o:1, n:0) later (p:0, o:1, n:0).
- Sentence 2: Reading (p:0, o:1, n:0) this (n/a) book (p:0, o:1,n:0) is (n/a) happy (p:0.875, o:0.125, n:0).

A word whose sentiment value is the greatest in positive, negative, or objective orientation is defined as a positive, negative, or objective word, respectively.

Jasmine Bhaskar, Sruthi K.,Prema Nedungadi proposed an enhanced technique for sentiment classification of online reviews by considering the objective words [5] and intensifiers[6].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. The polarity of the sentence can be obtained by following equation.

Sentence Score= $\Sigma I=1$ Score(i)

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.

M. Govindarajan proposed new hybrid classification method is proposed based on coupling classification methods using arcing classifier and their performances are analyzed in terms of accuracy.[7] A Classifier ensemble was designed using Naïve Bayes (NB), Support Vector Machine (SVM). In the proposed work, a comparative study of the effectiveness of ensemble technique is made for sentiment classification. The ensemble framework is applied to sentiment classification tasks, with the aim of efficiently integrating different feature sets and classification algorithms to synthesize a more accurate classification procedure.[7]

Muhammad Faheem Khan, Aurangzeb Khan and Khairullah Khan proposed a new method of word sense disambiguation (WSD) using matrix map of the semantic scores extracted from SentiWordNet of WordNet glosses terms.[8] The correct sense of the target word is extracted and determined for which the similarity between WordNet gloss and context matrix is greatest. Experiment results have shown that the proposed method improves the result of sentence level sentiment classification as evaluated on different domain datasets. From the result it is clear that the propose method achieves an accuracy of 90.71% at sentence level sentiment classification of online reviews.[8].

III. SENTIWORDNET AND WORDNET

SentiWordNet is sentiment analysis lexical resource made up of synset from WordNet, a thesaurus-like resource; they are allocated a sentiment score of positive, negative or objective. These scores are automatically generated using the semisupervised method which is described in [6]. It is also available freely for research purpose on web. SentiWordNet is one of the sources of sentiment analyses. It is a semiautomatic way of providing word/term level information on sentiment polarity by utilizing WordNet database of English terms and relations. WordNet is is a very rich source of lexical knowledge Since most entries have multiple senses. Each term in WordNet database is assigned a score of 0 to 1 in SentiWordNet which indicates its polarity. Strong partiality information terms are assigned with higher scores whereas less bias/subjective terms carry low scores. SentiWordNet is made up of a semi-supervised method which refers to a subset of seed terms to obtain semantic polarity. Each set of synonymous terms is assigned with three numerical scores ranging from 0 to 1 which indicates its objectiveness i.e. positive and negative bias [10]. One of the key features of SentiWordNet is that it assigns both positive and negative scores for a given term according to the following rule [6]:

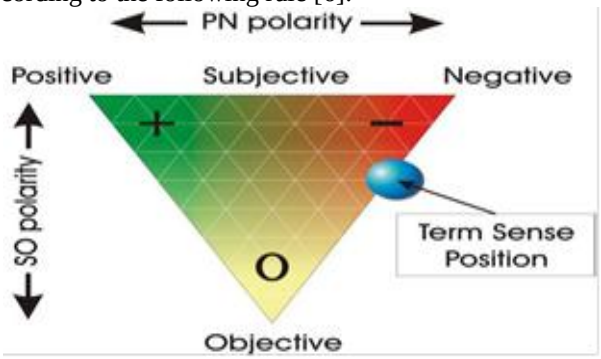

Fig. 1: Graphical Representation for SentiWordNet

- Pos(s) Positive score for synsets.
- Neg(s) Negative score for synsets.

– Obj(s) Objectiveness scores for synsets.

Then the following scoring rule applies:

Pos(s) + Neg(s) + Obj(s) = 1

The positive and negative scores are always given, and objectiveness can be implied by the relation:

Obj(s) = 1 – (Pos(s) + Neg(s))

Polarity scores according to synset and relevant part of speech are grouped by SentiWordNet database as a text file.

IV. PROPOSED METHODOLOGY

In this work, we propose an enhanced technique for sentiment classification of online reviews by considering the objective words. The proposed work consists of three modules document pre-processing, modifications of objective words in SentiWordNet and sentiment classification.

A. Data Pre-Processing Module Consists Following Steps

Preprocessing follows the same step as traditional text mining which consists of sentence splitter, POS tagging 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.

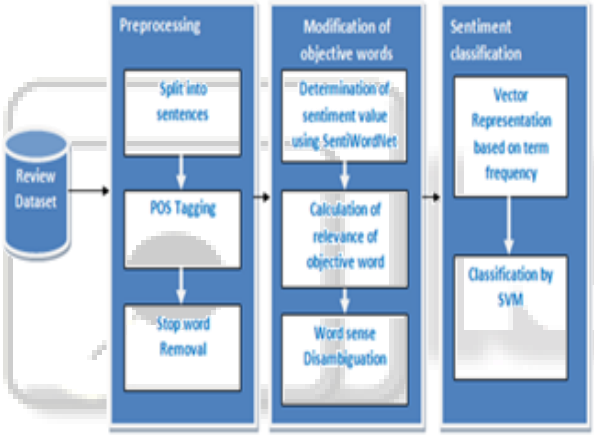

Fig. 2: Graphical representation for proposed method

1) Split Document into Sentences

Review document consists of several sentences and each sentence expresses the specific emotion. Reviewing the documents such as a comma, question mark, exclamation point or a period, depending on the punctuation as many sentences are first divided.

2) Part Of Speech (POS) Tagging

We use pos tagger to assign tag to each word. SentiWordNet provides four POS : 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.

3) Stop Word Removal

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 pre-processing.

B. Modification Of Objective Word In Sentiwordnet Module Consists Following Steps:

In this module calculate the relevance of an objective word and its associated sentences. The basic concept is that a positive or a negative sentence has some sentimental influence on its associated objective words. A positive sentence contains greater positive value and usually has more positive words than negative words. Thus, a positive sentiment tag is assigned to an objective word when this word appears in a positive sentence more often than in a negative sentence, and vice versa.

1) Sentiwordnet For Positive, Negative Score And Objective Score

Each word in the list updated with Positive and Negative score from SentiWordNet lexical resource. One of the key features of SentiWordNet is that it assigns both positive and negative scores for a given term according to the following rule [5,6]: For a synset s, we define.

– Pos(s) Positive score for synsets.
– Neg(s) Negative scores for synsets.
– Obj(s) Objectiveness scores for synsets.

Then the following scoring rule applies:

Pos(s) + Neg(s) + Obj(s) = 1;[5,6]

The positive and negative scores are always given, and objectiveness Can be implied by the relation:

Obj(s) = 1 – (Pos(s) +Neg(s))[5,6]

2) Algorithm for reassigning new value to objective words [5]

If (Objective word occurs only in positive sentence)then
PosWi=Psi/fri; NegWi=0; ObjWi=1-PosWi;
Else if(Objective word occurs only in
negative sentence)then
NegWi =Nsi/fri; PosWi =0;
ObjWi=1- NegWi;
Else if (Occurrence of Positive sentence< Occurrence of Negative sentence) then
If (Nsi-Psi>threshold) then
NegWi =Nsi/fri; PosWi =0; ObjWi=1- NegWi;
End if
Else if (Occurrence of Positive sentence>Occurrence of Negative sentence) then
If (Psi-Nsi>threshold)then
PosWi=Psi/fri;NegWi=0; ObjWi=1-PosWi;
End if
Else
PosWi=0;NegWi=0;ObjWi=1;
End if

3) Wordnet For Word Sense Disambiguation

The resultant score of the sentence are still not perfect because it has ambiguity as the word of natural language can be used in different senses having different scores according to the context. It is very difficult to decide automatically that which sense is used in the sentence. In this work new method is developed for correct sense extraction during sentiment analysis. Firstly senses of individual words are extracts from WordNet Glosses. To select the correct sense we create a matrix of similarity scores of word and senses of words.Finally if correct sense is extracted then will get the correct score of words from where the sentence score will be improved.To extract the true sense of the word in many senses of the word we present you with the number and the real sense of the word are the positive and negative scores use the WordNet glosses.

Algorithm for Word sense disambiguation [8]

1) INPUT:

Word_List:= All Sentiment Words in a sentence

AMB_Word:= Ambigues Word in a Word_List
WRD_GLOSSES_LIST: WordNet Glosses for AMB_Word
WRD_GlOSS: AMB_Word In WRD_GLOSSES_LIST
2) OUTPUT:
W_SENSE:= WordNet Sense for Word in Word_List
ForeachAMB_Word in Word_List
SELECT W_POSITIVE_SCORE AND W_NEGATIVE_SCORE FROM SentiWordNet
Foreach WRD_GLOSS in WRD_GLOSSES_LIST
SELECT G_POSITIVE_SCORE AND G_NEGATIVE_SCORE FROM SentiWordNet
If W_POSITIVE_SCORE AND W_NEGATIVE_SCORE is similar G_POSITIVE_SCORE AND G_NEGATIVE_SCORE
WRD_GLOSS then W_SESNSE:=WRD_GLOSS
End if
End for
End for

C. Sentiment Classification

1) Vector Representation Of The Document

SVM has high accuracy as compared to other machine learning approaches At first, represent the document as a vector Di = [WI, W2, W3 ... Wn] Where Wi is the weight of the term i with respect to the document.

2) Sentiment Classification Using SVM

Support vector machines are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM is a non-probabilistic binary classifier or binary linear classifier, i.e. it predicts, for each given input, which of two possible classes the input is a member of. Since an SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other[2]. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class, since in general the larger the margin the lower the generalization error of the classifier[2].

V. 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 for the products reviews of digital camera from which we have taken 314 sentences and we have used the data set taken from imdb.com for the moive reviews which we have taken 121 sentences for training and 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 given below.

Accuracy = TP+TN/TP + FP + TN + FN

From the result we can see that the prediction accuracy of the proposed method compare with SVM and NB in which SVM is much better than the NB. Reassigning objective word as positive or negative, improved the classification accuracy by reducing both positive and negative misclassification.. This is because miss-classification is less in the proposed method as compared to the existing method.

Dataset	Classifier	Accuracy
Movie Review Dataset	SVM	77.69
	NB	72.72
Product Review Dataset	SVM	81.97
	NB	74.32

Table 1:

VI. CONCLUSION

We have seen that sentiment analysis has many applications and it is important field to study. Sentiment analysis has Strong commercial interest because Companies want to know how their products are being perceived and also Prospective consumers want to know what existing users think. In this work proposed method used to improve the sentiment classification of product reviews by considering the objective words and handle word sense disambiguation problem combine. In this work we have used support vector machine and naïve bays for sentiment classification. From the result we can see that the prediction accuracy of the proposed method compare SVM and NB in which SVM is much better than the NB.

ACKNOWLEDGMENT

We take the immense pleasure in expressing our humble note of gratitude to our project guide, Ms Jayna Shah, Assistant Professor, Department of Computer Engineer,Sardar vallabhbhai Institute of Technology, for this remarkable guidance and suggestions, which helped us in completion of paper.

REFERENCES

[1] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10.Association for Computational Linguistics, 2002.
[2] en.wikipedia.org/wiki/
[3] Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." Proceedings of the 40th annual meeting on association for computational

linguistics. Association for Computational Linguistics, 2002.

[4] Yan Dang; Yulei Zhang; Hsin chun Chen, "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews, ―Intelligent Systems, IEEE , vol.25, no.4, pp.46,53, July-Aug.2010 doi: 10.1109/MIS.2009.105

[5] Hung, Chihli, and Hao-Kai Lin. "Using objective words in SentiWordNet to improve word-of-mouth sentiment classification." IEEE Intelligent Systems 28.2 (2013): 0047-54.

[6] Bhaskar, J.; Sruthi, K.; Nedungadi, P., "Enhanced sentiment analysis of informal textual communication in social media by considering objective words and intensifiers," Recent Advances and Innovations in Engineering (ICRAIE), 2014 , vol., no., pp.1,6, 9-11 May 2014 doi: 10.1109/ICRAIE.2014.6909220

[7] Govindrajan M. "Sentiment classification of movie review using hybrid method."

[8] Muhammad faheem Khan, Aurangzeb and khairullah khan―efficient word sense disambigution teqnique for sentence level sentiment classification of online review‖ Sci.Int(Lahore).25(4),2013

[9] Lee, Wei Jan, and Edwin Mit. "Word sense disambiguation by using domain knowledge." Semantic Technology and Information Retrieval (STAIR), 2011

