Reviews Classification Using SentiWordNet Lexicon

Alaa Hamouda Mohamed Rohaim
Department of Systems and Computers Engineering
Al_Azhar University

***Abstract*-Opinion lexicons are resources that associate sentiment polarity for words. SentiWordNet is one of these lexicons that assigns to each synset of WordNet three sentiment numerical scores, positivity, negativity and objectivity. Using opinion lexicons in opinion mining research stems from the hypothesis that individual words can be considered as a unit of opinion information, and therefore may provide clues to review sentiment. This paper presents the results of applying the SentiWordNet lexical resource using various techniques to the problem of automatic sentiment classification of reviews. Our approach yields an improvement over the previously used approaches for using SentiWordNet in sentiment classification of reviews. The results indicate SentiWordNet could be used as an important resource for sentiment classification tasks.**

I. INTRODUCTION

Opinion mining concerns with the reviews an author expresses. With the rapid growth of available subjective text on the internet in the form of product reviews, blog posts and comments in discussion forums, business analysts are turning their eyes on the internet in order to obtain factual as well as more subtle and subjective information (opinions) on companies and products. Opinion mining can assist in a number of potential applications in areas such as search engines, recommender systems and market research.

There are several approaches for detecting sentiment in text present in literature. One of the most important is to use the lexical resources such as a dictionary of opinionated terms. SentiWordNet[1] is one such resource, containing opinion information on terms extracted from the WordNet database - by using a semi supervised learning method - and made publicly available for research purposes. SentiWordNet provides a readily available database of term sentiment information for the English language, and could be used as a replacement to the process of manually deriving opinion lexicons.

In this paper we study various methods to classify product review as positive or negative using SetiWordNet lexicon. This study explores that SentiWordNet can be used in effective way to classify product reviews. The remainder of the paper is organized as follows: Section 2 presents the previous work of classifying products reviews. Section 3 describes the SentiWordNet lexicon and how it is used to make polarity classification, while section 4 presents our proposed technique to improve the performance of classification. The implementation and results are given in section. Finally, the conclusion and future work are presented in Section 6.

II. RELATED WORK

Classifying product reviews is a common problem in opinion mining and varieties of techniques have been used to address the problem. These techniques can be classified into two main approaches, as approaches based on lexical resources and neutral language processing and approaches employing machine learning algorithms.

Machine learning either supervised or unsupervised methods using different aspects of text as sources of features have been proposed in the literature. Early work seen in [2] presents several supervised learning algorithms using bag-of-words features common in text mining research, with best performance obtained using support vector machines in combination with unigrams. Classifying terms from a review into its grammatical roles, or parts of speech has also been explored: In [3] part of speech information is used as part of a feature set for performing sentiment classification on a data set of newswire articles, with similar approaches attempted in [4] and [5], on different data sets. In [6] part of speech, words string and root information are used with various combinations for performing classification on various data sets of consumer reviews. Separation of subjective and objective sentences for the purposes of improving review level sentiment classification are found in [7], where considerable improvements were obtained over a baseline word vector classifier. Other studies used lexical resources like SentiWordNet to build a data set of features derived from its scores to be used as features for support vector machines classifier as done in [8] and [9].

Opinion lexicons are resources that associate sentiment polarity for words. Their use in opinion mining research stems from the hypothesis that individual words can be considered as a unit of opinion information, and therefore may provide clues to review sentiment and subjectivity. In [8] SentiWordNet lexicon was applied by counting positive and negative terms found in a review and determining sentiment polarity based on which class received the highest score. The below section describes the SentiWordNet lexicon and how it was used in the previous work.

III. SENTIWORDNET

SentiWordNet [1] is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet [11] three sentiment numerical scores, positivity, negativity and objectivity, describing how Positive, Negative and Objective

the terms contained in the synset are. Each of the three scores ranges from 0.0 to 1.0, and their sum is 1.0 for each synset. This means that a synset may have nonzero scores for all the three categories, which would indicate that the corresponding terms have, in the sense indicated by the synset, each of the three opinion-related properties only to a certain degree. SentiWordNet word values have been semi-automatically computed based on the use the semi-supervised method described in [11]. Examples of sntiment scores associated to SentiWordNet entries are shown in Figure 1; the entries contain the parts of speech category of the displayed entry, its positivity, its negativity, and the list of synonyms. We show various synsets related to the words "good" and "bad". There are 4 senses of the noun "good", 21 senses of the adjective "good", and 2 senses of the adverb "good" in WordNet. There is one sense of the noun "bad", 14 senses of the adjective "bad", and 2 senses of the adverb "bad" inWordNet.

Category	WNT Number	pos	neg	Synonyms
A	01123148	0.875	0	good#1
A	00106020	0	0	good#2 full#6
A	01125429	0	0.625	bad#1
A	01510444	0.25	0.25	big#3 bad#2
N	03076708	0	0	trade_good#1 good#4 commodity#1
N	05144079	0	0.875	badness#1 bad#1

Figure (1): SentiWordNet Fragment

As described in [9] the Word sentiment interpretation was done by computing the number of times the 'word#sense' entry is more positive than negative (positive > negative), the number of times is more negative than positive (positive < negative) and the total number of entries 'word#sense' (or word) in SentiWordNet. Then the word is positive if it has more entries where the positive score is greater than negative score, negative if it has more entries where the negative score is greater than the positive score and neutral when there is no variation between positive and negative.

In [8] the word positive and negative scores were got from SentiWordNet according to its part-of-speech tag then applied a Term Counting method to classify review as positive or negative. In this method the lexicon was applied by counting positive and negative words found in a review and determining sentiment polarity based on which class received the highest score.

These techniques do not consider the positivity, negativity, and objectivity magnitude for each word. For example, both 'good' and 'acceptable' terms are counted as positive words regardless their positive magnitude. This limitation in these techniques led us to use other calculation approaches, as will be described in the next section.

IV. PROPOSED TECHNIQUE TO USE SENTIWORDNET IN REVIEW CLASSIFICATION

The aim of this work is to improve the efficiency of using the SentiWordNet in the task of sentiment classification for the reviews. The proposed technique consists of two phases. The first is to assign expressive (positive and negative) scores to the words; *SentiWordNet Interpretation phase*. The second phase applies various calculations on reviews to classify them as positive or negative using words scores got from the first phase; *Sentiment Polarity calculation phase*.

1. **SentiWordNet Interpretation Phase**: As mentioned in previous section each word has multiple senses, therefore to identify the positivity or negativity of a given word in text, one first needs to perform word sense disambiguation (WSD). In the current work we ignore WSD and proceed in the following way: for each word in SentiWordNet lexicon, positive and negative scores are calculated by getting the average for its entries according to category (Adjective, Adverb, Noun and Verb). For example a word "good" has 4 senses of the noun "good", 21 senses of the adjective "good", and 2 senses of the adverb "good", therefore it will has three positive and negative scores according to the three categories as shown in Figure 2.

Category	Average Pos Score	Average Neg. Score	Word
Noun	0.531	0	good
Adjective	0.5	0	good
Adverb	0.188	0	good

Figure (2): A word "good" average positive and negative scores

2. **Sentiment Polarity Phase**: As the Term Counting method did not take into consideration the magnitude value of positive and negative scores for words which have a significant effect on the level of word sentiment, therefore we consider the effect of magnitude value in the following calculations methods.

'Term Score Summation' method: in which the summation of positive and negative scores for each term found in a review, is calculated to get the positive and negative scores for all review words. Then, the review sentiment is determined based on which score has the highest value. This method has an advantage over Term Counting method that it takes into consideration the magnitude scores for words.

'Average on Sentence and Average on Review' method: for each sentence in the review, positive and negative scores determined by calculating the average of positive and negative scores for each term found in it. Then calculate the average of positive and negative scores for these sentences to get the positive and negative scores for the review, and determining sentiment polarity based on which score has the highest value.

Since each word in the SentiWordNet lexicon has three values for its positive, negative and objective sentiment and since the sum of these scores equal to 1, therefore the words which have low positive and negative scores have high objective scores. So, these words most probably have a neutral sentiment. Different thresholds applied to neglect the words which more objective score than positive or negative

ones. For example, threshold 0 means neglect words which have positive and negative scores equals to 0.

V. IMPLEMENTATION AND RESULTS

We implemented the proposed technique using the General Architecture for Text Engineering (GATE) tool, a framework for the development and deployment of language processing technology in large scale [12]. We used "Amazon Product Review Data Set" which is available at http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html. This data set has many reviews for different products (book, camera, mp3, etc.) and rated from 1 to 5 point. So, we converted it to binary classes (positive and negative) by considering the reviews with rates (1, 2) as negative reviews and the reviews with rates (4, 5) as positive reviews and neglected the reviews with rate 3. From this data set we selected 2000 reviews as positive and 2000 reviews as negative provided that their size is not more than 500 characters.

As a preparation phase before applying the proposed technique, we made the following linguistic analysis processes for reviews, figure 3:

- *Tokenization process*: splits the text into very simple tokens such as numbers, punctuation and words of different types.
- *Sentence Splitting process*: segments the text into sentences. This module is required for the tagger. The splitter uses a gazetteer list of abbreviations to help distinguish sentence-marking full stops from other kinds.
- *Speech Tagging process*: produces a part-of-speech tag as an annotation on each word or symbol.

After that, we applied the proposed technique of 'Sum on Review' and 'Average on Sentence and Average on Review'. To be able to compare the accuracy of our technique with the previous work, we applied 'Term Counting' technique on the same corpus of Amazon reviews. The following subsections show the results obtained in each technique.

5.1 Term Counting

Applying the technique used in [8], positive and negative scores for each review were calculated by counting positive and negative words, and then the sentiment polarity was determined by assigning the review to the class with the highest score. This method yielded an overall accuracy of 56.77%, with results detailed in the table 1.

Table (1): Term Count Results

Class	**Positive**	**Negative**
Predict Positive	1641	393
Predict Negative	564	1286
Total	2000	2000
Class Recall	80.35%	35.7 %
Class Precision	55.55 %	64.5%

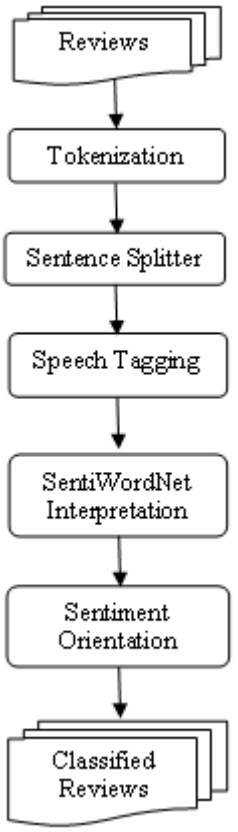

Figure (3): Sentiment Classification Phases

5.2 Sums on Review

Positive and negative scores for each review were calculated using 'Sum on Review' technique as mentioned in section 4 with various thresholds. This method yielded the results shown in table 2. It is apparent that the best accuracy was 67 % at threshold 0.

Table (2): Sums on Review Results

Threshold Value	Positive Reviews		Negative Reviews		Over All Accuracy
	Precision %	Recall %	Precision %	Recall %	
0.0	62.88	83.00	75.00	51.00	67.00
0.1	63.04	81.95	74.21	51.95	66.95
0.2	63.61	77.45	71.18	55.70	66.58
0.3	62.63	71.80	66.96	57.15	64.48

5.3 Average on Sentence and Average on Review

Positive and negative scores for each review were calculated using 'Average on Sentence and Average on Review' technique as mentioned in section 4 with various thresholds. This method yielded an overall accuracy of 68.63 % at threshold 0. The results are detailed in table 3.

Table (3): Average on Sentence and Average on Review Results

Threshold Value	Positive Reviews		Negative Reviews		Over All Accuracy
	Precision %	Recall %	Precision %	Recall %	
0.0	64.58	82.50	75.78	54.75	68.63
0.1	64.16	82.35	75.37	54.00	68.18
0.2	64.37	78.15	72.20	56.75	67.45
0.3	63.02	72.35	67.55	57.55	64.95

5.4 Comparison

The table below illustrates how the proposed methods to use SentiWordNet overcome the 'Term Counting' method in [8], where the same data set (Amazon reviews) were used in all experiments.

Table (4): Comparing Results

Method	Accuracy
Term Counting [8] applied on Amazon Corpus	56.77 %
Sum on Review (this research)	67.00 %
Average on Sentence and Average on Review (this research)	68.63 %

As shown in table, 'Average on Sentence and Average on Review' method yields an improving over the term count method, while using threshold to neglect words with high neutral scores decreases the accuracy.

VI. CONCLUSION AND FUTURE WORK

SentiWordNet is a valuable lexicon used to sentiment classification for reviews. The previous work of Term Counting technique produced an accuracy of 56.77%, while our proposed techniques which considered the sentiment magnitude of both words and the sentences produced significant improvement in the accuracy to be 67% and 68.63%. It is also noted that using threshold to neglect words which have high neutral scores does not improve the accuracy. As future work, we can make more improvement in the accuracy by considering negation expressions that exchange the polarity of the sentence sentiment.

REFERENCES

[1] A. Esuli and F. Sebastiani, SentiWordNet: *A Publicly Available Lexical Resource for Opinion Mining*. Proceedings from International Conference on Language Resources and Evaluation (LREC), Genoa, 2006.

[2] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? *Sentiment Classification using Machine Learning Techniques*. Proceedings of EMNLP, 2002.

[3] T. Wilson, J. Wiebe, and P. Hoffmann. *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis*. Proceedings of HLT/EMNLP, Vancouver, Canada, 2005.

[4] A. Kennedy and D. Inkpen. *Sentiment Classification of Movie Reviews Using Contextual Valence Shifters*. Computational Intelligence, Vol. 22, 110–125, 2006.

[5] F. Salvetti, S. Lewis and C. Reichenbach. *Automatic Opinion Polarity Classification of Movie Reviews*. Colorado Research in Linguistics. Volume 17, Issue 1 (June 2004). Boulder: University of Colorado.

[6] A. Funk, Y. Li and H. Saggi, K. Bontchevaon and C. Leibold. *Opinion Analysis for Business Intelligence Applications*. Proceedings of the first international workshop on Ontology-supported business intelligence, 2008.

[7] B. Pang and L. Lee. *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*. Proceedings of the ACL, 2004.

[8] B. Ohana and B. Tierney. *Sentiment Classification of Reviews Using SentiWordNet. 9th*. IT&T Conference, Dublin Institute of Technology, Dublin, Ireland, 22nd.-23rd. October, 2009.

[9] H. Saggionα and A. Funk. *Interpreting SentiWordNet for Opinion Classification*. Proceedings of the Seventh conference on International Language Resources and Evaluation LREC10, 2010.

[10] C. Fellbaum. WordNet - *An Electronic Lexical Database*. MIT Press, 1998.

[11] A. Esuli and F. Sebastiani. *Determining term subjectivity and term orientation for opinion mining. In Proceedings of EACL-06*, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT. Forthcoming, 2006.

[12] Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. *GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications*. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), 2002.