Sentimental Analysis of Book Reviews using Unsupervised Semantic Orientation and Supervised Machine Learning Approaches

Vipin Deep Kaur
Netaji Subhas Institute of Technology
New Delhi, INDIA
vipindeepkaur25@gmail.com

Abstract—Sentimental analysis aims at identifying the opinions of various users. This paper presents my research work on the application of sentimental analysis on book reviews. I have applied both unsupervised (Semantic Orientation - Pointwise Mutual Information - Information Retrieval) and supervised (Support Vector Machine and Naïve Bayes) machine learning approaches on two openly available book review datasets from GoodReads and Amazon. The comparative analysis of the approaches on the datasets indicates that unsupervised approach performs better on GoodReads dataset with an accuracy of 73.23% whereas supervised approach gives better results on Amazon dataset with Naïve Bayes giving the maximum accuracy which ranges from 73.72% to 74.73% in the case of 5-folds and 10-folds respectively.

Keywords: Book reviews, Sentimental analysis, Support Vector Machine, Naïve Bayes, Unsupervised approach

I. INTRODUCTION

The advent of internet and technology has facilitated the users with a higher access to web applications through smart devices and mobile phones thus improving product rating system immensely. Now, a customer can become an active user by giving reviews about different products/services which may be useful to other potential customers. But, there are hundreds, thousands or even more product/service related reviews available on the web and reading all those available reviews is a very tedious and taxing task for the customer [1]. Therefore, there is a need gap for apt techniques which automatically summarize these reviews into a positive or a negative category to give useful information to the user. This task of classification of reviews by identifying the opinions of various users is formally known as Opinion Mining or Sentimental Analysis [2].

Sentimental analysis may be defined as the classification of a text or document into a positive or a negative class by judging the connotation contained in the text. A positive opinion expressing text is assigned a positive label whereas a negative label denotes a negative opinion [1]. Any objective opinion would be assigned a neutral label.

It is observed that significant work has been done in the domain of product reviews [4], [5], movie reviews [1], [2], [7], restaurant reviews [14], blog posts [7] etc. to identify their sentiments but comparatively very less work has been done in the domain of book reviews [5], [8]. Hence, this paper targets sentimental analysis in book domain.

Researchers have explored various sentimental analysis techniques such as:-

i) Supervised approaches like Support Vector Machine (SVM) [1], [2], [4], [5], [8], Naïve Bayes (NB) [1], [2], [4], [5], Random Forest (RF) [5], Maximum Entropy (ME) [4] etc. and ii) Unsupervised approaches like Semantic Orientation - Pointwise Mutual Information - Information Retrieval (SO-PMI-IR) [1], [2], [3], [6], [9], [10], SentiWordNet (SWN) [2], [7] etc.

Among the above-mentioned techniques, I have chosen SO-PMI-IR technique which computes the polarity of reviews by extracting the opinionated words from the reviews using Part-of-Speech (POS) tagging, evaluating their Semantic Orientations (SO) and then aggregating these SO scores to decide the overall class of the review [9]. I have taken two datasets from GoodReads [11] and Amazon book reviews [12].

Further, the results are compared with NB and SVM techniques. These techniques are the most popular ones for sentimental analysis [13].

The results show that unsupervised approach performs better on GoodReads dataset whereas supervised approach gives better results on Amazon book reviews with NB giving the maximum accuracy.

The rest of the paper has been organized as follows. Section II consists of literature review and the methodologies used have been described in Section III. The experimental setup has been presented in Section IV whereas Section V presents the results. Section VI includes the conclusion and future work.

978-1-5386-5657-0/18/$31.00 ©2018 IEEE

II. LITERATURE REVIEW

Several algorithms have been applied in the field of sentimental analysis over the past few years [13]. Peter D. Turney [9] proposed an unsupervised approach of SO-PMI-IR to categorize reviews as thumbs-up (positive) or thumbs-down (negative). The observed accuracy varied from 66% for movie reviews to 84% for automobile reviews. V K Singh et al [3] also used SO-PMI-IR to mine the students' opinion regarding different subjects by collecting feedback from them in textual format.

P Walia et al [1] explored unsupervised (SO-PMI-IR) as well as supervised approach (NB and SVM) for sentimental analysis of movie reviews. The results showed that SO-PMI-IR gave the best accuracy and NB outperformed SVM. V K Singh et al [2] also explored SWN technique along with NB, SVM, and SO-PMI-IR on movie reviews.

Xing Fang and Justin Zhan [5] proposed a new feature vector generation algorithm to perform sentiment polarity categorization of product reviews (beauty, books, home, and electronics) obtained from amazon.com.

T. K. Shivaprasad and J. Shetty [4] presented the taxonomy of various sentimental analysis algorithms. They explored NB, SVM and ME based supervised approaches on reviews from sports, electronics, and computer.

Rodrigo Moraes et al [8] studied the performance of SVM and artificial neural network (ANN) on book dataset and mentioned the scope of sentimental analysis in the book domain.

The related work discussed above has been summarized in table I. As evident from the table, NB, SVM, and SO-PMI-IR are the most promising techniques in the field of sentimental analysis.

III. METHODOLOGIES USED

The complete process of sentimental analysis followed in this research has been shown in fig 1.

As shown in fig 1, the first step is dataset preparation in which blank and unreadable reviews are removed from the dataset and then processed reviews are extracted to text files. Then, in the second step, two supervised approaches namely, NB and SVM and an unsupervised approach namely, SO-PMI-IR have been used for sentiment classification which are discussed below.

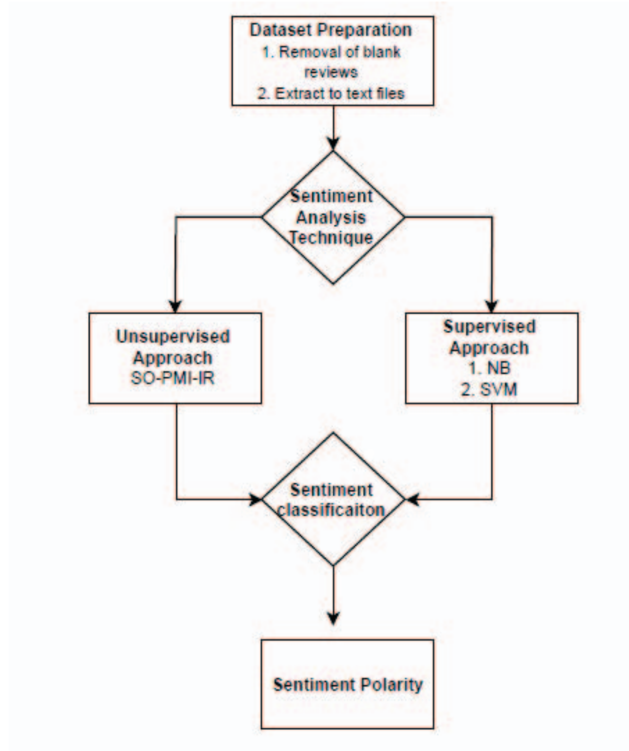

Fig 1. Sentimental analysis process

A. SUPERVISED APPROACH

A supervised learning approach is the process of an algorithm learning from a training data and then performing the required classification on the test dataset.

1. Naive Bayes Approach

NB classifies the text statistically. It can be applied to perform sentimental categorization of textual reviews into one of the two classes, either positive or negative. Deciding whether some particular kind of words will express opinions more concretely or all the words in the textual review should be considered as features like in a normal classification problem [1] is an important issue here. Naïve Bayes is a type of Bayes theorem probabilistic learning method as given in eq 1.

$$P(C/D) \alpha P(C) \Pi_{1\leq k \leq nd} P(t_k/C) \qquad (1)$$

The statistical pattern in which the words/terms occur in a document helps in text categorization. These selected words are often termed as features. Most of the researchers suggested that adjectives or adverbs are good sentiments expressing words, therefore, selecting words with these tags

TABLE I. SUMMARY OF RELATED WORK

S. No.	Title of the paper	Dataset used	Approach used	Results
1	P. Waila et al [1]	Three movie review datasets	SVM, NB, and SO-PMI-IR	NB gives better accuracy than SVM and similar to SO-PMI-IR.
2	V. K. Singh et al [2]	Two datasets from Cornell polarity dataset and third created by own from Hindi films	SVM, NB, SentiWordNet (SWN), SO-PMI-IR	NB performs comparably to SVM. SO-PMI-IR gives high accuracy and seems the best choice, but it takes a lot of time to compute Pointwise Mutual Information (PMI) values. SWN can be used as it requires no training dataset.
3	V. K. Singh et al [3]	Created own dataset containing around 1000 student reviews	SO-PMI-IR	Largely accurate.
4	T. K. Shivaprasad and J. Shetty [4]	Sports, electronics, and computer reviews	NB, SVM, and ME	SVM gives the highest accuracy and NB and ME gives medium accuracy
5	Xing Fang and Justin Zhan [5]	Product reviews from Amazon	NB, SVM, and Random forest	RF performs better on datasets either on manually-labeled or machine-labeled sentences.
6	V. K. Singh et al [7]	Two existing movie dataset from Cornell. Blog posts on Libya & Tunisian revolution	Applied different variants of SWN	The SWN technique achieves commendable performance levels. Also, it does not require any prior training or training data unlike the other supervised techniques.
7	Rodrigo Moraes et al [8]	GPS, books and camera reviews from Amazon	SVM and artificial neural network (ANN)	ANN gives better accuracy on all datasets but SVM remains unaffected by noisy terms in misbalanced datasets.
8	Peter D. Turney [9]	Dataset from Epinions.com	SO-PMI-IR	Algorithm attains an average accuracy of 66% on movie dataset and 84% on automobile dataset.
9	H. Kaur et al [13]		Study of various sentiment analysis process, categorization techniques	Sentimental analysis and classification have a great scope and NB and SVM are the two most popular techniques for sentimental analysis.

could be a good choice of features for classifying documents into negative or positive classes [9].

Bernoulli and multinomial NB are two popular type of NB approaches. Multinomial NB not only considers the absence or presence of a word/term in the document but also takes into account the number of times it is present in the document as an indicator for a specific class. Whereas Bernoulli NB does not consider the frequency of a word in the document, it only considers the absence or presence of the word. I have applied multinomial NB.

2. Support Vector Machine

SVM is a classifier based on vector space model which converts strings/documents into feature vectors before classification. It tries to find the largest margin between two classes. The aim is to find an optimal hyperplane which is a dividing region between two classes that is away from all the training elements (text documents in this case) as far as possible. A margin of the classifier is evaluated by the distance of the nearest data point from the decision surface [8]. These divider points are called as support vectors. More is the margin, less is the uncertainty in the classification of classes. This maximization is achieved by SVM.

Fig 2 shows an SVM classification model [15]. Here, '+' and 'o' are two separate classes. A, B, and C are hyperplanes. Since the distance of all the training points is largest from A thus A provides the best separation.

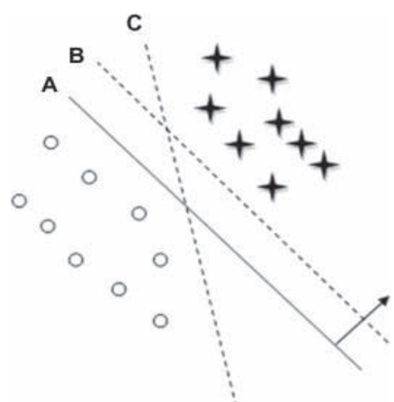

Fig 2. SVM classification method [15]

In this paper, the text documents are converted to multidimensional term frequency and inverted document frequency *(tf.idf)* vectors.

B. UNSUPERVISED APPROACH

An unsupervised approach overcomes the disadvantage of the supervised approach as it does not need any prior labeling of training dataset to perform the classification. I have applied SO-PMI-IR technique for classifying reviews. It consists of three steps: -

i) Extraction of phrases using *pos (part-of-speech) tagging:* -All the words in a document do not express opinions, therefore, only opinion expressing words are extracted. For this work, I have used adjectives [1].

ii) Calculating the semantic orientation of these phrases: - The SO of all the extracted phrases within a document is calculated using PMI as given in the eq 2.

$$\mathrm{PMI}(\mathrm{Phrase}_1, \mathrm{Phrase}_2)=\log\{\mathrm{Prob}(\mathrm{Phrase}_1 \Delta \mathrm{Phrase}_2) / \mathrm{Prob}(\mathrm{Phrase}_1).\mathrm{Prob}(\mathrm{Phrase}_2)\} \quad (2)$$

Here *Prob*($Phrase_1$ Δ $Phrase_2$) provides the co-occurrence probability of the two phrases.

The *Prob*($Phrase_1$).*Prob*($Phrase_2$) computes the probability that two phrases co-occur if they are statistically independent.

The ratio between *Prob*($Phrase_1$ Δ $Phrase_2$) *and Prob*($Phrase_1$).*Prob*($Phrase_2$) measures the statistical independence between these two phrases.

Moreover, the log of this ratio computes the amount of information about the presence of phrase1 when we observe phrase2 or vice versa [9]. Thus, SO of a complete phrase can be evaluated as illustrated in the eq. 3,

$$\mathrm{SO}\ (\mathrm{Phrase}) = \mathrm{PMI}(\mathrm{Phrase}, \text{“excellent”}) - \mathrm{PMI}(\mathrm{Phrase}, \text{“poor”}) \quad (3)$$

Where, PMI(phrase, "excellent") gives the relationship of the phrase with excellent (a positive standard word) and PMI(phrase, "poor") gives the relationship of the phrase with poor (a negative standard word) [9]. Negation of an adjective (words like not bad, not interesting etc.) is handled by negating the semantic orientation of an adjective preceded by a 'not' (E.g. if SO value of excellent is 0.67 then if preceded by not it becomes -0.67). I have calculated SO scores of phrases using python.

iii) Classifying the class of the document: -This step calculates the overall SO of the review by taking the aggregation of the SO scores of individual phrases making up the document. Aggregation scheme used can be a sum, max, min or any other function. A review is assigned a positive or a negative label by comparing the aggregated SO value with a fixed threshold. I have used sum as an aggregate function. If the sum is greater than 0, the review is assigned a positive label otherwise a negative label.

IV. EXPERIMENTAL SETUP

This section describes the datasets used and the performance measures of the research.

A. Dataset Description

Two datasets containing book reviews from GoodReads and Amazon have been used.

Following the standard customer 5-rating system [3], review with a rating < 3 is called a negative review, one with a rating = 3 is called a neutral review and that with a rating > 3 is called a positive review. Even after the pre-processing of raw data, the data obtained contains an unevenly high number of positive reviews. So, an equal number of both types of reviews are taken to avoid the bias of the classifier towards positive reviews in case of supervised approach. Finally, GoodReads dataset contains a total of 2746 book reviews and Amazon dataset contains a total of 6832 reviews both consisting of an equal number of positive and negative labeled reviews.

B. Performance Measures

The following measures have been used to analyse the performance of the algorithms used:-

Accuracy = (True Positive + True Negative) / Total instances
Precision = True Positive / (True Positive + False Positive)
Recall = True Positive / (True Positive + False Negative)
F-score = 2. (Precision. Recall/ (Precision + Recall))

I have also performed 5- fold and 10 –fold cross-validation to further avoid bias while learning.
The techniques used in this paper have been implemented in Python.

V. RESULTS

This section shows the detailed discussion of the results obtained. I conducted experiments on both datasets. Table II shows the computed results of the NB, SVM and SO-PMI-IR approaches on GoodReads reviews and Table III shows the computed results of these techniques on the Amazon book reviews.

TABLE II. RESULTS OF THE ALGORITHMS ON GOODREADS DATASET

Method	Performance measure	5-fold	10-fold
NB	Accuracy	67.4%	68.31%
	Precision	0.708	0.719
	Recall	0.591	0.600
	F-score	0.644	0.654
SVM	Accuracy	69.84%	70.02%
	Precision	0.694	0.692
	Recall	0.709	0.720
	F-score	0.701	0.705
SO-PMI-IR	Accuracy	73.23 %	
	Precision	0.821	
	Recall	0.806	
	F-score	0.813	

The unsupervised approach gives better results than supervised approach on GoodReads dataset with an average accuracy of 73.23%.

TABLE III. RESULTS OF THE ALGORITHMS ON AMAZON BOOK REVIEWS

Method	Performance measure	5-fold	10-fold
NB	Accuracy	73.72%	74.73%
	Precision	0.729	0.741
	Recall	0.754	0.759
	F-score	0.741	0.749
SVM	Accuracy	73.46%	74.70%
	Precision	0.726	0.739
	Recall	0.752	0.762
	F-score	0.738	0.750
SO-PMI-IR	Accuracy	62.30 %	
	Precision	0.645	
	Recall	0.545	
	F-score	0.59	

The average accuracy of NB ranges from 73.72 % to 74.73% in the case of 5 and 10 folds and that of SVM is 73.46 % when 5-fold is applied and 74.70% when 10-fold is applied.

The following figure shows the accuracy comparison for three approaches on two datasets where dataset 1 is GoodReads dataset and dataset 2 is Amazon dataset.

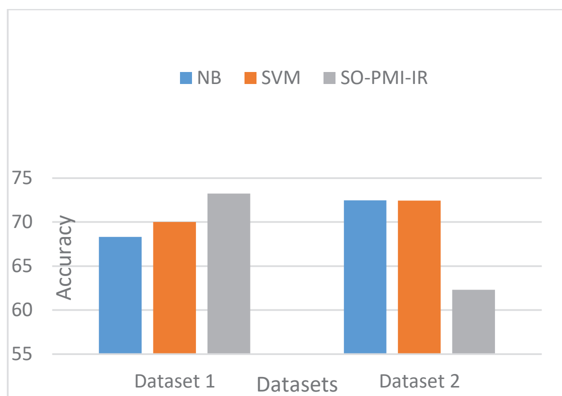

Fig 3. Accuracy values for the three approaches on two datasets

The unsupervised approach did not give satisfactory results on Amazon book reviews; the reason for this may be due to the fact that it contains short one-lined phrases thus few opinions containing words might be present for extraction.

VI. CONCLUSION AND FUTURE WORK

This work focuses on the sentimental analysis of book reviews using both supervised and unsupervised approaches. For this purpose, I have applied the popularly used techniques namely NB, SVM and SO-PMI-IR on two datasets from GoodReads and Amazon. The results show that unsupervised algorithms gave better results when the dataset contains long phrases whereas supervised algorithms give higher accuracy on the dataset containing short one-lined reviews.

Adjectives have been used in this research. Adverbs are the words that modify the adjectives, therefore, extraction of adverbs along with adjectives or using some other pattern of phrase extraction remains a future work. Also, further pre-processing of the dataset may also help.

REFERENCES

[1] P. Waila, Marisha, V. K. Singh and M. K. Singh, "Evaluating Machine Learning and Unsupervised Semantic Orientation approaches for sentiment analysis of textual reviews," *2012 IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, 2012, pp. 1-6.

[2] V. K. Singh, R. Piryani, A. Uddin, P. Waila and Marisha, "Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches," *2013 5th International Conference on Knowledge and Smart Technology (KST)*, Chonburi, Thailand, 2013, pp. 122-127.

[3] V. K. Singh, P. Kumari, A. Singh and J. Thapa, "An automated course feedback system using opinion mining," *2011 World Congress on Information and Communication Technologies*, Mumbai, 2011, pp. 72-76.

[4] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, 2017, pp. 298-301.

[5] Xing Fang and Justin Zhan, "Sentiment Analysis using product review data," Springer: Journal of Big data, North Carolina A& T State University, Greensboro, NC, USA, 2015.

[6] P. Chaovalit and L. Zhou, "Movie Review Mining: Comparison between Supervised and Unsupervised Approaches," In *Proceedingsofthe38th Hawaii International Conference on System Sciences*, 2005.

[7] V. K. Singh, R. Piryani, A. Uddin and P. Waila, "Sentiment analysis of Movie reviews and Blog posts," *2013 3rd IEEE International Advance Computing Conference (IACC)*, Ghaziabad, 2013, pp. 893-898.

[8] Rodrigo Moraes, João Francisco Valiati, Wilson P.Gavião Neto. "Document-level sentiment classification: An empirical comparison between SVM and ANN." *Expert Systems with Applications* 40.2 (2013): 621-633.

[9] Peter D. Turney," Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the *40th Annual Meeting of the Association for Computational Linguistics (ACL),* Philadelphia, July 2002, pp. 417-424.

[10] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 79-86, Philadelphia, US, 2002.

[11]https://www.kaggle.com/gnanesh/goodreads-book-review

[12] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. "Image-based recommendations on styles and substitutes." *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 43-52. ACM, 2015.

[13] H. Kaur, V. Mangat and Nidhi, "A survey of sentiment analysis techniques," *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, 2017, pp. 921-925.

[14] Hanhoon Kang, Seong Joon, and Yoo Dongil Han., "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Syst Appl*, 39:6000
10, 2012.

[15] Walaa Medhat, Ahmed Hassan and Hoda Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, 5(4), 1093–1113, 2014.