International Journal on Emerging Technologies **11**(1): 376-386(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

A Novel Approach for Creating Sentiment Lexicon for Movie Reviews using Star Ratings

Shashank Sharma[1] and Gautam Dutta[2]
[1]Research Scholar, Indian Institute of Foreign Trade, Delhi, India.
[2]Professor, Indian Institute of Foreign Trade, Delhi, India.

(Corresponding author: Shashank Sharma)
(Received 14 October 2019, Revised 16 December 2019, Accepted 24 December 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Lexicon based methods employ popular sentiment lexicons like MPQA and SentiWordNet which are built using the principle of Pointwise Mutual Information (PMI) to calculate the Sentiment Orientation (SO) of words. These lexicons are generally used for sentiment classification tasks like polarity determination of reviews, but their performances are mostly inferior to supervised methods. While lexicon methods are easier to implement and can generalize better across datasets, the key challenge lies in building lexicons that promise performance on par with supervised methods. The popular PMI based methods used for creating lexicons do not make use of the rich information that can be obtained from the label or any other such information associated with text documents like reviews. For example, the rating points or star ratings that accompany the reviews can provide the polarity intensity of reviews. Star ratings can be used to generate SO for key words by using concept of conditional probabilities which has been used in some of the recent researches to develop sentiment lexicons. This paper contributes to this growing body of work and proposes a novel approach (SentiDraw) where the probability distribution of words across reviews with different star ratings is used to calculate their SO scores. Lexicons created using this method were tested with contemporary movie reviews across two sub-domains (Hollywood and Bollywood) along two other popular movie review datasets for comparison and benchmarking of methods. The results showed that SentiDraw lexicons deliver state-of-the-art performance which is better than any other lexicon-based method to the knowledge of the authors. When compared to performance of popular lexicons like SentiWordNet, the accuracy obtained is more than 25% better along with significant improvement in F-scores as well.

Keywords: Classification of reviews; lexicon; movie reviews; polarity determination; SentiDraw; sentiment analysis.

Abbreviations: SA, sentiment analysis; ML, machine learning; PMI, pointwise mutual information; SVM, support vector machine.

I. INTRODUCTION

A range of efforts aimed at decoding consumer reaction for a given product and service have attempted to utilize the fingerprints left by the consumers in the digital media. A particularly useful 'fingerprint' is the digital text that is getting generated at a phenomenal rate everyday by users across the world on various digital platforms. In case of several product categories like movies, a huge amount of electronic word of mouth (eWOM) gets generated every time a new movie gets released. More and more consumers use online discussion forums, consumer review sites, weblogs, social network sites etc. to exchange product information [1]. Various researches have established the high correlation between online reviews and product sales [2-4]. The effect of online reviews is also much more exaggerated in case of experience goods like movies which have a very short life and one doesn't have much insight into how good or bad her transaction is going to turn out before he or she makes the transaction and watches the movie [5]. One must make the purchase going not by his or her own evaluation but mostly on evaluation of others [2].
Several studies have also established the high degree of causative role that online reviews and comments play on the outcome of a movie [3, 6-10].
However, given the abundance of eWOM like reviews, tweets and comments that get generated in a very short time after a movie's release it is impractical to manually label the reviews based on the sentiment polarity of the opinions (positive or negative) and predict the success or failure of movie in the required time to maximize its utility. In recent years though, natural language processing based supervised and unsupervised methods have been shown to be quite efficient at this labeling task [11]. The automation makes it practical to quickly sieve through millions of reviews and determine the polarity (positive or negative) of individual reviews to predict the success of the movie. This automation for extracting sentiment expressed in a text is called Sentiment Analysis (SA) and determination of polarity is a specific task within SA called Sentiment Classification (SC). One of the goals of the studies in the domain of sentiment analysis of movie reviews is to increase the accuracy of sentiment classification based on input data, mostly in the form of text.

Many of the NLP methods of feature extraction and engineering have been coupled with a host of Machine Learning (ML) techniques to increase accuracy of SA. Some of the examples of such algorithms are Logistic Classification [12], Naïve Bayes [13,14], Support Vector Machines [13-15], and Rule based Classifiers [16]. The major issue with supervised learning is the availability of labeled corpora and computational complexity [13]. Another approach for sentiment classification makes use of existing or built-for-purpose sentiment lexicons that provide a sentiment orientation score for many words and phrases using which the overall sentiment polarity or even the sentiment strength of the reviews can be calculated [17]. The latter have the advantage of avoiding the hard-working step of labeling training data. However, these techniques rely on (external) lexical resources which are concerned with mapping words to a categorical (positive, negative, neutral) or numerical sentiment score, which is used by the algorithm to obtain the overall sentiment conveyed by the text. Therefore, the effectiveness of the whole approach strongly depends on the goodness of the lexical resource it relies on [17].
Several generic sentiment lexicons like SentiWordNet [18], MPQA Subjectivity Lexicon [19], Linguistic Inquiry and Word Count (LIWC) [20], SentiStrength [21], SentiWords [22], Affective norms for English word (ANEW) [23] and General Inquirer [24] have been created over the last few years which have been popularly used in sentiment analysis tasks. Lexicons can be created manually [25], or automatically, using seed words to expand the list of words [26-28]. Most studies use the concept of PMI (Pointwise Mutual Information) to create lexicon from a set of seed words [24]. Some of these lexicon-based studies have used only adjectives as predictors of the semantic orientation of text [26]. In these researches, all adjectives are extracted and annotated with their SO value, using the dictionary scores. The SO scores are in turn aggregated using different scoring methods into a single score for the whole text [17].
Using sentiment lexicon for sentiment analysis is convenient since they are much faster and less computationally intensive compared to ML based methods. Also, they don't require training data with polarity labels or trained models to begin with [11]. In terms of performance, ML based sentiment analysis gives better result in most cases, but these models often perform poorly when used on a different domain [22]. Even for unsupervised classification using lexicons, building domain specific sentiment lexicons have been shown to increase the performance significantly [44]. While some researchers have attempted to build a domain specific lexicon from scratch using various methods [29], others have sought to modify existing lexicons for a given domain [30]. There is performance enhancement in both cases but not enough to deliver result on par with ML methods. Recent studies have used concept of conditional probabilities by calculating sentiment scores based on probability distribution of words across positive and neutral documents [45, 46]. The lexicons obtained from these methods have been shown to perform much better in sentiment classification tasks compared to PMI based method. This study has contributed to this approach by enhancing the technique further by making use of probability distribution of key words across all rating points of labelled reviews to build the lexicon and then testing the method with both contemporary movie reviews across two sub-domains (Hollywood and Bollywood) and two other popular movie review datasets for comparison and bench marking. The main contributions of this research paper are a) to propose a novel method of building state-of-art sentiment lexicon for polarity determination of reviews which delivers a significantly better performance than any other lexicon developed earlier and can be treated on par with ML methods in terms of performance, b) to use latest reviews across two sub-domains of movie reviews (Hollywood and Bollywood) in order to understand the impact sub-domain specificity on accuracy performance and c) to compare results of proposed lexicon with other commonly used lexicons on two of the most popular movie data sets used across various studies for benchmarking.
The lexicon created by this approach has been named SentiDraw as it determines the Sentiment orientation score of each word by using its probability Distribution across star Ratings for all selected Words. These words, along with their sentiment scores, are then used to determine the polarity of reviews. The accuracy scores delivered by this sentiment lexicon range from 82% to 90% depending on the domain and has a significantly better performance on even the most experimented dataset like Cornell movie reviews data [13] and Large movie review data set [31] compared to any other lexicon based technique to the knowledge of the authors. The next section highlights some key contributions in the space of sentiment classification. There is more emphasis on polarity determination studies done for movie reviews to enable a better comparison of the method used in this paper as the dataset chosen for comparison belongs to the domain of movie reviews. However, the methods used in the reviewed papers are not limited to movie reviews only and have a general applicability across all domains where reviews are available along with their ratings for polarity classification task. The third section details the methodology and explains the SentiDraw algorithm developed for sentiment scoring. It also describes the corpus used for building the SentiDraw sentiment lexicon. The fourth and fifth sections present the results and discuss the findings respectively.

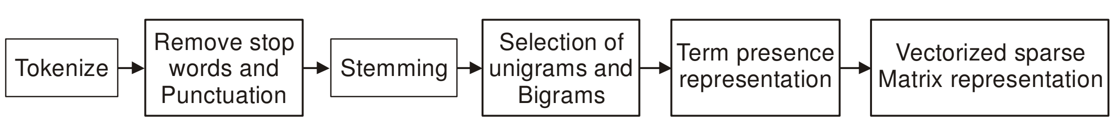

Fig. 1. General Steps of feature processing and representation for supervised algorithms.

II. RELATED WORK

One of the earliest landmark papers on polarity classification of movie reviews by [14] used a data set of 1400 Hollywood movie reviews from IMDB. IMDB is a popular movie database website and one of the most widely consulted movie review website for user reviews and ratings. They used SVM algorithm with unigrams as features and could achieve accuracy of 86% with their methodology. Consequently, Pang and Lee [13] improved on this method using minimum cuts and used Cornell movie reviews dataset which was a dataset of 1400 Hollywood movie reviews from IMDB to compare performance of the results. Several papers published since then have used the same Cornell movie reviews dataset for polarity classification study. Among the various algorithms used in these studies, Support Vector Machine (SVM), Naïve Bayes (NB) and Maximum Entropy (ME) are the most favored [32].

The first step for polarity classification task begins with data cleaning. The data is generally in the form of text which must be prepared so that it can be used for model learning and classification task. The unstructured text needs to be modified into a structured format that can be further processed in meaningful manner. In data mining in general and text mining specifically the pre-processing phase is of significant impact on the overall outcomes [28]. This step is important not only to put the data in a structured format but also to select and represent the features in the most optimum manner.

Bag of words (BoW) is the most commonly used feature in which the words or phrases are simply represented as a multiset based on their presence in a document and the frequency disregarding grammar. Pang and Lee [13] used bag of features framework with unigrams alone with using both unigrams and bigrams together and found unigrams alone to be slightly more accurate. While term frequency – inverse document frequency (TF-IDF) had been very successful in other domains like topic classification, it was shown [13] that frequency of words may not be a good predictor for sentiment. Instead binary representation of each feature (presence or absence) can lead to more accurate prediction.

Other than supervised ML based methods, other approaches make use of part of speech information with sentiment orientations of the words. They are used either stand alone or used along with ML methods to improve accuracy. Opinion words are words that are commonly used to express positive or negative sentiments [34]. For example, adjectives like nice, fantastic, good, and great are positive opinion words while ugly, bad, and boring are negative opinion words. Semantic Orientation (SO) is a real number measure of the positive or negative sentiment expressed by a word of phrase [26]. Mullen and Collier [35] selected and defined the SO scores of the adjectives using 'Point Mutual Information' method with the seed words like "excellent" and "poor" and then denoting a SO value to each objective using below formula for calculating PMI. In these methods a set of seed words are used for which sentiment orientation is already known. Using these seed words, other words are allocated a sentiment orientation based on their probability of co-occurrence with these seed words.

For any given corpus (generic or domain-specific) or web searches, the difference of probability of co-occurrence of a given word with both positive and negative oriented seed words give the sentiment orientation scores of the word. Researches employ different scoring methods for this. While some allocate continuous scores ranging from -1 to +1 where positive scores indicate positive orientation and negative scores indicate negative orientation, others may give discrete values. SentiWordNet [18] provides information about polarity identification as well as for subjectivity detection. It provides discrete values for sentiment score where 0.25, 0.5, 0.75 and 1.0 show positive polarity with varying degree of positivity with 1 being highest. The case is exactly reverse in case of negative scores. Using sentiment lexicons, the sentiment orientation scores can be calculated at a phrase level, sentence level or document level. Common way to aggregate the scores at any level is to average out the sentiment score values of all the words present in the text. If the final score is positive, the text is considered as positively oriented or else negative if the average score comes out to be negative [17].

Various other methods of building domain specific lexicons have also been employed. Sentiment lexicons can be generated (1) manually; (2) using a dictionary; or (3) using a corpus of documents. The most common approach of using PMI-IR method was described above [26]. Another method explored [19] presented a lexicon-based method for sentiment analysis of text where a sematic orientation calculator (SO-CAL) was developed for the detection of sentiment orientation using dictionaries of words annotated with their semantic orientation (polarity and strength. Esuli and Sebastiani [36] exploited the glosses information from Wordnet. Qiu *et al.*, [37] adopted dependency relations between sentiment words and aspect words. Wiebe [38] utilized the dependency triples from an existing parser. Hu and Liu [49] used the synonym and antonym relations within linguistic resources. Most of the current approaches study the adaptation or sentiment transfer learning of a trained classifier (supervised techniques) or lexicon (unsupervised techniques) from one domain to another which involves having a general lexicon to start with, but very few works actually focus on techniques that build specific domain lexicons without requiring a-priori knowledge [48]. Similar work by Thelwall [21] developed a lexicon called SentiStrength using a lexicon of 2310 sentiment words and word stems obtained from the LIWC program [20] and General Inquirer [22] and optimized its lexicon term weights for a specific set of human-labeled texts. It does this by repeatedly increasing or decreasing the term weights by 1, one term at a time, and then assessing whether this change increases, decreases or does not affect the overall classification accuracy for the human coded texts. A domain specific sentiment lexicon produces better sentiment detection results as compared to a general-purpose lexicon like SentiWordNet [18]. A new domain specific sentiment lexicon, named SentiCircles, was proposed by Saif *et al.,* [29] which provides domain dependent prior polarities for sentiment detection. The prior polarities were updated with respect to the context.

Significant improvement in accuracy and F-measure was observed when comparison of results was performed with SentiStrength [21] and SentiWordNet [18]. Another method of creating a domain specific lexicon using information retrieval is to assign a score for each word based on the difference between the probabilities of its occurrence in positive text and negative text used by Labille *et al*., [48]. This method was also employed by Lee *et al*., [47] and was shown to achieve an F score of 89% for opinion words when compared to human coders. A trade-off between precision and coverage is hard to find. Gatti *et al*., built SentiWords, that has both a high precision and a high coverage. It blends SentiWordNet with newer lexicons in a learning framework using an ensemble method to create a prior polarity lexicon of approximately 155,000 words [22].

Several researchers have compared the performance of these lexicons on text classifications tasks such as polarity determination. Kim *et al*., [39] compared alternative supervised learning methods including SO-PMI, conditional probability of words or polarity, and simply frequency-based method using movie review data set from IMDB (Internet Movie Data Base) which included 80,000 movie reviews. They found that lexicons using conditional probability of words employing term frequency-based methods show better accuracy than that using SO-PMI. A new general-purpose sentiment lexicon called WKWSCI Sentiment Lexicon was developed and compared with five existing lexicons: Hu & Liu Opinion Lexicon, MPQA, General Inquirer, NRC Word-Sentiment Association Lexicon and SO-CAL40. WKWSCI, MPQA, Hu & Liu and SO-CAL were found to be equally good for product review sentiment categorization, obtaining accuracy rates of 75% to 77%. A benchmark comparison of twenty-four popular sentiment analysis methods was performed by Ribeiro *et al*., [41] who identified the top nine methods for 2-class classification task based on Macro-F. These were SentiStrength, Sentiment140, Semantria, Opinion Lexicon, LIWC15, SO-CAL, AFINN, VADER and Umigon. Some other advanced methods of creating sentiment lexicon have also been explored recently which try to emulate human learning to ensure continuous improvement in the robustness of the lexicon [42, 43].

As the simplest method used for classification, use of n-grams with any of the ML classification methods among SVM, NB and ME deliver a decent performance with accuracies above 80% and reaching as high as 86% with SVM14. These baseline methods can be used as benchmarks to explore other methods of sentiment classification using different approaches including use of sentiment lexicon. So far, to the knowledge of the authors, no sentiment lexicon based methods, even domain specific ones, have been able to match the performance of ML based methods. Table 1 and 2 list some of most cited studies done on the domain of movie reviews using supervised and lexicon based methods respectively along with the classification performance for a comparison of methods that can a serve as benchmark for the method proposed in this paper.

Table 1: Summary of classification method and results using supervised methods.

Classification Method	Results		Year	Reference
NB/SVM	NB: 86.4%	SVM: 87.1%	2004	[13]
SVM	90.20%	—	2005	[51]
RBC-SBC-GIBC-SVM	F-score for SVM- 87.30	—	2008	[16]
Logistic Classification	92.70%	—	2011	[12]
SVM	10 folds: 91.7%	Bootstrap: 91.52%	2010	[52]
Ensemble Combining NB, ME, SVM	Joint-POS and metaclassifer: 86.85%	WR-based feature sets and metaclassifier: 88%	2011	[53]
SVM	F test= 71.89%	—	2013	[54]
SVM and ANN	SVM: 85.2% with 1000 features	ANN: 86.5% with 3000 features and 86% with 1000 features	2013	[55]
Expected Sentiment Weight (ESW)	86.04% accuracy	—	2014	[56]

Just like earlier findings based on comparative evaluation of lexicons [38], the best results obtained among studies in Table 2 show that lexicons [45, 46] using some form of conditional probability basis the label of the reviews or corpus deliver best performance in classification tasks. Senti-CS [45] builds on SentiWordNet3.0 by using it as a labeled corpus. The part of speech information, usage-based ranks and sentiment scores are used to calculate Chi-Square-based feature weight for each unique subjective term/part-of-speech pair extracted from SentiWordNet3.0. This weight is then normalized in a range of -1 to +1 using min–max normalization. Another method [46] which is quite close to method proposed in this paper generates the lexicon directly from any labeled corpus for any language without the need to start with the small set of words as a seed or any existing lexicon. It defines the classification buckets basis the labels and then used conditional probability of occurrence of each word in each of the bucket to give it a sentiment score.

Lexicon obtained from this method was tested on movie review dataset and obtained a high performance with F-score of 82%. Recently similar methods have been employed for classification of posts on social media as well [47, 48] by generating a domain-specific lexicon using probabilities and information theoretic techniques. In method employed in these papers, the probabilistic score Scoreprob (w) of a word w is computed using posterior probabilities and is defined as the difference of the probability of w of being positive, p(pos |w), and its probability of being negative. The novel state-of-the-art method proposed in this paper builds on this approach and improves it substantially by introducing star rating dispersion for each word. The accuracy and F-scores obtained are greater than 85% for the two sub-domains of Hollywood and Bollywood movie reviews selected for experiment in this paper which is the best performance for any lexicon-based method to the knowledge of the authors.

Table 2: Summary of classification method and results using lexicon based methods.

Lexicon used	Accuracy	Year	Reference
Reverse JST (T30)	69.1%	2010	[57]
JST (T30)	70.2%	2011	[58]
Lexicon Labelling	66.9%	2011	[59]
GI Lexicon	75.0%	2012	[60]
SentiWordNet (APS)	65.0%	2013	[61]
HM Lexicon	55.0%	2014	[62]
SentiMi	75.0%	2015	[29]
SentiWordNet	72.1%	2016	[20]
Term Frequency and Bootstrapping	70.2%	2016	[63]
Senti-CS	~80%	2016	[45]
SPLM	82.0%	2017	[46]

III. MATERIALS AND METHODS

A. Sentidraw Framework

This paper demonstrates a method of creating sentiment lexicon that is comparable in performance with ML based methods and out performs any other sentiment score methods in at least domain specific polarity classification. The novel state-of-the-art method SentiDraw uses star ratings of the reviews and the probability distribution of these star ratings across words for calculating sentiment scores. The method builds on the work of [35, 38, 45-49] but advances the method considerably by integrating star ratings instead of only polarity of the reviews. Also, the scores aren't calculated using simply the difference in probability as done in most methods. The initial scores obtained by the probabilities are weighted using a scoring schematic based on reviews and further normalized to give final sentiment orientation scores.

Reviews of various product categories are available on sites like Amazon, IMDB, Yelp, Zomato and others. These reviews are almost always labeled with a star rating that is often on a scale of 5 or 10 with higher star ratings indicating higher intensity of positive sentiment. SentiDraw method makes use of the star ratings that accompany the reviews to calculate the SO scores for each key word.

In the first step, each review was first tokenized and tokens were removed if they belonged to stop words or were a punctuation mark. Each word was then stemmed using Porter stemming algorithm [64] and subsequently POS-tagged using Stanford-POS tagger [65]. Python was used for these steps as it contains useful libraries in its nltk package for carrying out these tasks. Among the words only the nouns, adjectives, adverbs and verbs were selected along with their POS tags these tokens were named following SentiWordNet norm [18]. SentiWordNet also only contains words that belong to one of these four parts of speech: adjective, noun, adverb and verb as these parts of speech often contain a lot more subjectivity than any other parts of speech [24]. These parts of speech are represented respectively as 'a', 'n', 'r', 'v'. So a word like 'good' is selected and named as 'good.a' or 'good.r' depending on its usage as adjective or adverb as determined by the POS tagger.

The total count of each token 't' across all reviews in the corpus was denoted as C_t in Also the frequency of these tokens across all rating points for the selected reviews in the selected corpus were counted and we denoted it as $f_{t,r}$ for each token 't' and rating 'r'. A dictionary was created with the token and its frequency count for each star rating. The probability of a given token for each rating point is then calculated by dividing its frequency count for a rating point by its total frequency count across all rating points.

$$P(t,r) = \frac{f_{t,r}}{C_t} \quad (1)$$

Next a weighted average sentiment score is calculated for each word by using its probability of occurrence at each rating point and a prior determined sentiment score for each rating point. IMDB uses a 10-scale rating and the movie reviews selected for this research were also taken from IMDB. A sentiment score between '-5'and '+5' was then ascribed to each rating point like the rating scheme used by [19] where rating points higher than '5' were given positive scores to connote positive sentiment and the value of this score was higher for higher ratings. Rating points '5' and below '5' were scored negatively with decreasing value for lower ratings. This scoring patter is since most research papers studying movie reviews consider '5' and '6' as neutral ratings for the review [32]. So low value of -1 and +1 for these rating points are close to neutral score of 0 and yet having a small negative and positive value respectively to differentiate between these two rating points. Table 3 shows the sentiment score for each rating point:

Table 3: Sentiment Scores for Rating Points.

Star Rating	Rating Sentiment Score (R_s)
1 and 2	-5
3 and 4	-3
5	-1
6	+1
7 and 8	+3
9 and 10	+5

The weighted average sentiment scores (WS_t) for each token 't' is calculated by:

$$WS_t = \sum_{i=o}^{r} P(t, r_i) \times R_{s,i} \quad (2)$$

The weighted average sentiment scores thus determined for the words are further normalized using two different approaches and these approaches have also been compared basis their performance in this paper. The first approach is based on Standard Deviation and uses a two-step process. Sentiment Orientation scores (SO) for each word were calculated using the number of standard deviations (s) from the mean given the normal distribution of weighted average sentiment scores across tokens. Any score above +1 is rounded off to +1 and any score below -1 is rounded off to -1 to restrict the SO scores between -1 and +1 and to not allow outliers to influence the overall document level sentiment scores in a substantial manner.

$$SO_t = \frac{WS_t}{s} \quad (3)$$

Another approach to normalize the sentiment score between -1 and 1 is to apply the commonly used method called MinMax normalization where each individual score is divided by the difference between maximum and the minimum SO score obtained for the words that belong to the dataset under investigation.

$$SO_t = \frac{WS_t}{WS_{t,\,Max} - WS_{t,\,Min}} \quad (4)$$

This method is sometimes sensitive to outliers in the dataset. However, the below graph shows the distribution of sentiment scores in the dataset and it shows a smooth distribution making MinMax normalization method a worthy candidate to explore for normalization. Only a couple of data points in both sets had slightly extreme values towards the negative side and were excluded from MinMax normalization calculation.

B. Experiments

Dataset Description: IMDB was used for creating a corpus of movie reviews used in this study for developing the SENTIDRAW lexicon. IMDB (Internet Movie Database) is one of the most popular websites where hundreds of audience reviews can be found for most of the movies released across the world. This source of movie reviews has been used across several key researches on sentiment analysis of movie reviews [13, 14, 35, 38, 58, 59, 61-65]. For this study, a total of 83,500 reviews were extracted from IMDB for both Bollywood and Hollywood reviews for movies released between the years 2012 and 2017 using Python library 'Scrapy' for web scraping.

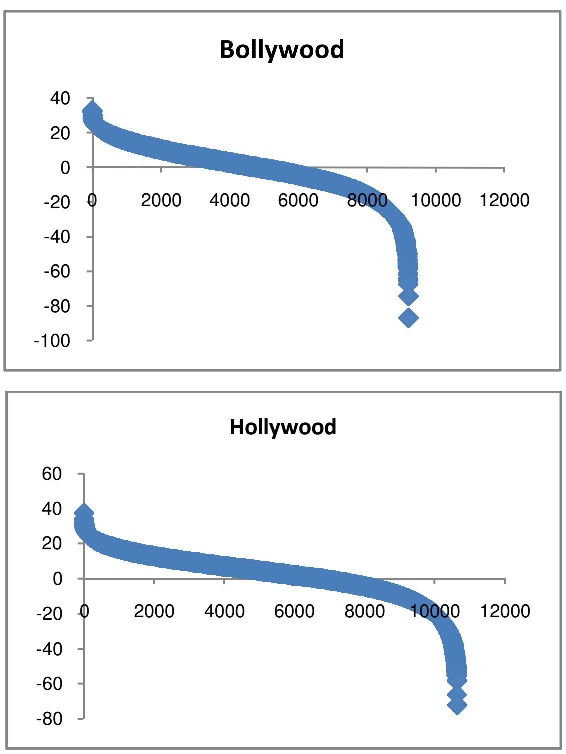

Fig 2. Distribution of terms basis polarity without normalization in SentiDraw Lexicons.

Many of the reviews were found to contain less than 5 words and were dropped from the sample set. Also, some of the reviews did not contain accompanying rating point and had to be dropped as well as they could not have been labeled.

Finally, 20,000 reviews each for Bollywood and Hollywood were selected for experiment. Further tests were done on the popular Cornell movie reviews dataset [13] and large movie review dataset [31] as these datasets have also been used in several other landmark studies applying different methods for polarity classification. Comparing performance on these datasets provides a good benchmark for comparing SentiDraw with other classification methods.

Methodology: The reviews extracted from IMDB are divided into training set and testing set so that we can test the lexicon on reviews that have not been used to create the lexicon enabling us to test if the lexicon can be generalized. The reviews in the training set are then processed by first tokenizing the reviews after which the stop words and punctuations are removed.

Stop words are words like 'a', 'the', 'in' etc. which commonly occur and do not really add to any information sentiment polarity of the text. The tokens are then stemmed using Porter stemmer and are tagged using Stanford POS-tagger. As discussed earlier, only adjectives, nouns, adverbs and verbs are then chosen as terms that will be used to build the SentiDraw sentiment lexicon. The SentiDraw lexicon is created the SentiDraw framework for building lexicon described above for each of the corpus (Bollywood reviews and Hollywood reviews).

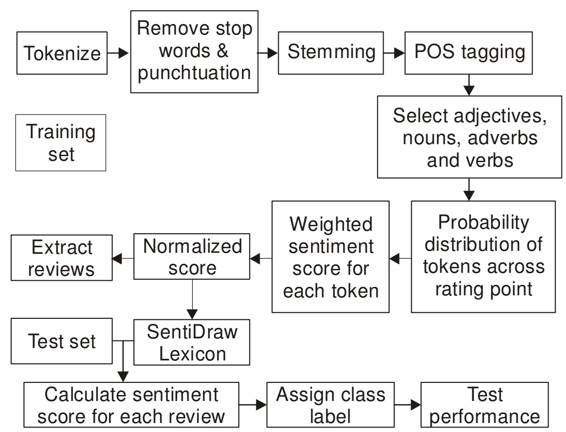

Fig 3. Overview of key steps for creating SentiDraw.

Using the SentiDraw method, sentiment lexicons lexicon and classifying reviews basis sentiment polarity were created for both Bollywood and Hollywood reviews using 15,000 reviews in each case. The sentiment scores of each review in the test data was calculate the average SO score using the SO value of tokens in the reviews. This was tested on 5,000 reviews each for Bollywood and Hollywood. Further tests were done on popular movie review data sets which have been used in several other studies using different methods so provide a good benchmark for comparison. SentiWordNet (SNW) is the most used sentiment lexicons. SNW's performance was also tested on all these test data sets and compared to SentiDraw.

Evaluation Metrics: Given the task of polarity determination consists of predicting either of the two classes for the samples, the performance of the algorithm can be evaluated basis commonly used metrics: accuracy, recall and precision. These metrics are calculated basis below given confusion matrix which represents a grid of actual classes versus predicted classes:

Table 4: Confusion Matrix.

	Predicted	
	Positive documents	Negative documents
Actual positive documents	No. of True Positive samples (TP)	No. of False Negative samples (TP)
Actual negative documents	No. of False Positive samples (TP)	No. of True Negative samples (TP)

The value of the three metrics is then given as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

Most studies declare only the 'Accuracy'. However, this metric is not very reliable in a case where there is a high skew towards any of the two classes. Although this is not generally the case for movie reviews, F-score has also been proposed as a more balance metric where both metrics 'Recall' and 'Precision' are combined as their harmonic mean:

$$\text{F-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

IV. RESULTS AND DISCUSSION

Fig. 4 shows how the ratings were dispersed across the selected 40,000 movie reviews for each of the sub-domains. It can be clearly seen that Bollywood reviews tend to have a much more skewed rating with almost 40% reviews rated as either '1' or '10' whereas Hollywood reviews have less skew with only 22% reviews rated as either '1' or '10'.

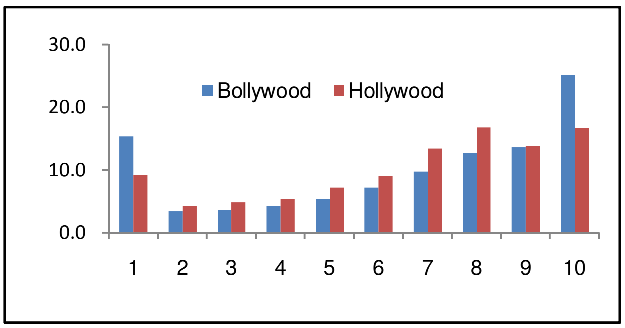

Fig. 4. Rating Dispersion of 20,000 movie reviews each for Bollywood and Hollywood.

Whether it is due to the nature of Bollywood movies themselves that they tend to be either too good or too bad or due to a more objective or nuanced opinion of Hollywood movie reviewers.

It can be concluded on the basis of extremity of ratings that Bollywood reviews are more extreme. This may have an impact of how much more accurately the algorithms are able to classify the polarity of Bollywood movie reviews versus Hollywood movie reviews.

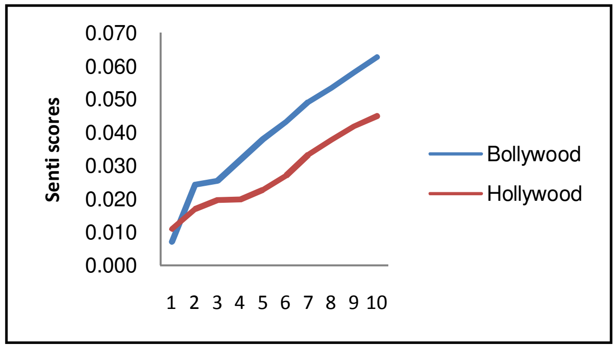

Fig. 5. Average sentiment scores across rating.

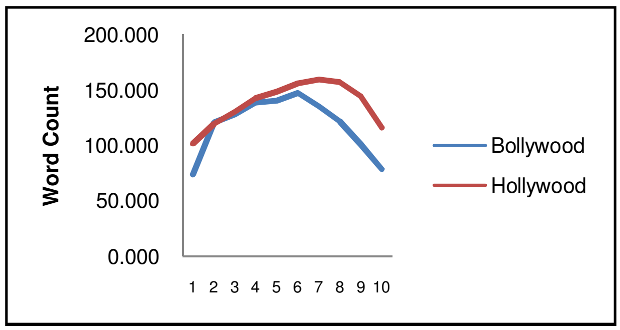

Fig. 6. Number of words per review across rating.

Both Hollywood and Bollywood lexicons were built using SentiDraw method with 15,000 reviews. Table 5 describes the distribution of terms based on their sentiment polarity along with their dispersion across parts of speech. As discussed earlier, two approaches for normalizing the sentiment scores of individual words have been used for creating the SentiDraw lexicons. Lexicons built with the two domains using both approaches are also compared below along with SentiWordNet basis their classification performance in their respective domains.

Table 5:SentiDraw Term Distribution based on polarity and their parts of speech.

	Verb		Noun		Adjective		Adverb	
	Positive	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**
Hollywood Reviews	1301	1076	2785	2201	1538	1060	433	355
Bollywood Reviews	1107	909	2455	1907	1207	963	207	328

After normalization of sentiment scores using both standard deviation method (SentiDraw STD) and MinMax method (SentiDrawMinMax), the scores for each review in the test dataset was calculated and compared with Sentiscore method. Table 6 compares the scores across both datasets using all the key measures.

Using SentiDraw lexicons for predicting sentiment polarity of movie reviews is clearly superior to SentiWordNet performance across all measures. The accuracy and F score obtained surpass any other lexicon-based method when it comes to accuracy and other key measures like recall, precision and F score, as per the knowledge of the authors. Accuracy scores of 87.5% and 89.5% for Bollywood dataset and that of 82.2% and 85.4% for Hollywood dataset using STD and MinMax normalization respectively illustrate state-of-the-art performance and are a considerable improvement over any existing method of creating lexicons in terms of domain specific accuracy for classification task. In fact, they are comparable to machine learning methods without the computational requirements of the ML techniques and requirement for labeled training set each time. The MinMax method of normalization also emerges as the better of the two normalization methods. It has a clear edge over STD method across measures except in the case of precision for positive reviews and recall for negative reviews.

It is evident that both SentiDraw lexicons obtained for Hollywood and Bollywood are quite an improvement over existing method in terms of classification performance when used on their respective datasets. However, it is important to understand their usage can be generalized to different movie review datasets which either do not belong to the same period or are taken from a different source. For this a comparison was done with the well accepted SentiWordNet lexicon on popular movie review datasets Cornel movie reviews dataset and Large movie reviews dataset which have been used across several studies and can be taken as a benchmark data for comparison. The Tables 7 and 8 show the relative performance of SentiWordNet lexicon with both SentiDraw lexicons of Hollywood and Bollywood using MinMax normalization on both benchmark datasets respectively. The results above show that SentiWordNet lexicon performs a lot poorly on both datasets when compared to SentiDraw. Both SentiDraw lexicons perform far better on accuracy scores and are also superior in performance compared to any other lexicon to the knowledge of the authors. This is a considerable improvement over current method.

Table 6: Comparison of classification performance.

	Bollywood			Hollywood		
Measures	**SentiWordNet**	**SentiDraw STD**	**SentiDraw MinMax**	**SentiWordNet**	**SentiDraw STD**	**SentiDraw MinMax**
Accuracy	66.6%	87.4%	89.5%	66.2%	82.2%	85.4%
Pos F-Score	72.5%	90.8%	92.5%	74.2%	86.5%	89.5%
Pos Recall	64.7%	90.1%	93.8%	69.8%	82.6%	90.1%
Pos Precision	82.4%	91.4%	91.2%	79.1%	90.7%	88.8%
Neg F-Score	57.3%	80.3%	82.8%	51.2%	73.9%	76.2%
Neg Recall	70.4%	81.6%	80.3%	58.0%	81.2%	75.0%
Neg Precision	48.2%	79.1%	85.6%	45.9%	67.9%	77.5%

Table 7: Comparison of classification performance on Cornell Movie Reviews Data Set [13].

	Cornell Movie Reviews Data Set [13]		
	Senti-WordNet	**SentiDraw**	
		Min Max (Holly)	**Min Max (Bolly)**
Accuracy	61.6%	79.7%	75.7%
Pos F-Score	54.5%	81.3%	72.7%
Pos Recall	66.9%	75.5%	83.1%
Pos Precision	46.0%	88.0%	64.6%
Neg F-Score	66.8%	77.9%	78.1%
Neg Recall	77.3%	71.4%	86.9%
Neg Precision	58.9%	85.6%	71.0%

Table 8: Comparison of classification performance on Large movie reviews dataset [31].

	Large Movie Reviews Data Set [31]		
	Senti-WordNet	**SentiDraw**	
		Min Max (Holly)	**Min Max (Bolly)**
Accuracy	56.9%	82.9%	81.5%
Pos F-Score	69.0%	83.4%	82.0%
Pos Recall	95.8%	85.5%	84.4%
Pos Precision	53.9%	81.3%	79.8%
Neg F-Score	29.4%	82.5%	81.0%
Neg Recall	17.9%	80.4%	78.6%
Neg Precision	81.0%	84.7%	83.4%

Domain specificity also has a clear role as can be concluded from the results above. The Hollywood SentiDraw lexicon performs better than the Bollywood SentiDraw lexicon since both these movie datasets are for Hollywood reviews so there is more similarity in domain. However, Bollywood lexicon is still significantly better than SentiWordNet and other lexicons in terms of performance even for Hollywood reviews which illustrate the generalizability of SentiDraw method.

V. CONCLUSION

Earlier studies had never used rating dispersion to arrive at sentiment scores for words. This intuitively appears to be an effective means to score the sentiment scores of words as words with more positive connotation will be likely to appear in reviews with better ratings and vice versa. Also, movie reviews on different platforms are often accompanied with rating on different scales and provides enough a ready supply of data to build lexicon for movie reviews using the rating dispersion. This paper attempts the same and builds lexicon for two different movie domains, Hollywood and Bollywood, to also study the extent to which domain specificity contributes as a factor in overall performance. The lexicons are created for each of the datasets, Hollywood and Bollywood, using 15,000 reviews from IMDB website using the 10-scale star rating used on the website and then scoring each word based on the dispersion of its co-appearance across each rating point. This lexicon, SentiDraw has been shown to outperform any lexicon-based method for sentiment classification task in the domain of movie reviews. The performance of this domain specific lexicon is far ahead of a more generic but widely used sentiment lexicon SentiWordNet. This clearly illustrates that the method of creating sentiment lexicon for reviews using the rating dispersion is a state-of-the-art method for creating sentiment lexicons. Also, lexicons have always been known to be domains sensitive and this study further illustrates the same. While the intrinsic nature of Bollywood reviews with use of more sentiment-oriented words with more skewed ratings allows for higher classification accuracy, the Hollywood reviews are more nuanced and classification performance for Hollywood reviews trails the former as result. However, the classification performance is still significantly high compared to other lexicon-based methods and is very close to that of ML based methods which are significantly more computationally intensive. Also, ML based methods have a poor record in performance when tested on different domains and Lexicons based methods tend to do better in a general scenario. The results obtained on Pang and Lee dataset illustrate this point. While Hollywood SentiDraw lexicon has better performance, even the SentiDraw built from Bollywood reviews obtained in last few years can classify benchmark movie reviews to 76%-82% accuracy levels. This performance is significantly better (by more than 25% in terms of accuracy) than that of the more commonly used SentiWordNet.

While the method is shown to have a lot of promise, the lexicon can benefit from several refinements which can be taken up in future studies. The named entities in the reviews, like the name of the movie itself, actor or director may not be very relevant for generalizing the reviews. Such words can be removed from the lexicon. Also, the scaling of the sentiment orientation scores is done basis a skewed sample set where positive reviews were twice as many as negative reviews. The performance may benefit from updating the scores basis a more balanced dataset. SentiWordNet is a more generic lexicon and widely used. Performance of the SentiDraw dataset can be compared with SentiWordNet across different domains to make an estimate of its cross-domain performance versus SentiWordNet. It can also be directly compared other popular lexicons like So-CAL, MPQA, SentiStrength, SentiWordNet and LIWC. The SentiDraw method employed above also did not use negation for sentences with negative words. This has been show to further refine sentiment scores and is worth attempting to compare the results. Another area of refinement stems from the need to normalize the scores. While the results above have shown that MinMax normalization works better than Standard deviation method, there are several other normalization techniques which can be explored in future studies. Lastly, Word Sense Disambiguation also helps in identifying the meaning of the word in each context. If word sense disambiguation is employed at the time of developing lexicon and used when the sentiment scoring is being done on a given text, the performance can improve considerably. Web based interface is now a very easy platform to collect online reviews and opinions about anything [65] and domain specific lexicons developed for decoding sentiments more accurately can enhance predictive power of eWOM significantly.

REFERENCES

[1]. Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, *21*(4), 23-45.

[2]. Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing, 70*(3), 74-89.

[3]. Elliott, C, & Ghosh Dastidar, S. (2019). The Indian film industry in a changing international market. *Journal of Cultural Economics. https://doi.org/10.1007/s10824-019-09351-6*

[4]. Godes, D., & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, *23*, 545-560.

[5]. Reinstein, D. A., & Snyder, C. M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The journal of industrial economics, 53*(1), 27-51.

[6]. Eliashberg, J., & Shugan, S. (1997). Film critics: Influencers or predictors? *Journal of Marketing*, *61*, 68-78.

[7]. Dellarocas, C., Awad, N., & Zhang, X. (2004). Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning. ICIS 2004 Proceedings, 1-34.

[8]. Duan, W., Gu, B., & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of retailing*, *84*(2), 233-242.

[9]. Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? The effect of tweets on movie sales. *Decision Support Systems*, *55*(4), 863-870.

[10]. Niraj, R., & Singh, J. (2015). Impact of user-generated and professional critics reviews on Bollywood movie

success. *Australasian Marketing Journal (AMJ),* 23(3), 179-187.
[11]. Waila, P., Singh, V. K., & Singh, M. K. (2012). Evaluating machine learning and unsupervised semantic orientation approaches for sentiment analysis of textual reviews. In 2012 IEEE International Conference on Computational Intelligence and Computing Research (pp. 1-6). IEEE.
[12]. Bai, X. (2011). Predicting consumer sentiments from online text. Decision Support Systems, 50(4), 732-742.
[13]. Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 1-8.
[14]. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-10 (pp. 79-86). Association for Computational Linguistics.
[15]. Abbasi, A., France, S., Zhang, Z., & Chen, H. (2011). Selecting attributes for sentiment classification using feature relation networks. *Transactions on Knowledge and Data Engineering*, *23*(3), 447-462.
[16]. Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), 143-157.
[17]. Musto, C., Semeraro, G., & Polignano, M. (2014, December). A Comparison of Lexicon-based Approaches for Sentiment Analysis of Microblog Posts. *Information Filtering and Retrieval*, 59-68.
[18]. Baccianella, S., Esuli, A., &Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec, *10*, 2200-2204.
[19]. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.
[20]. Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, *71*(2001), 2001.
[21]. Thelwall, M. (2017). The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength (summary book chapter). Cyberemotions: Collective emotions in cyberspace. Berlin, Germany: Springer, 119-134.
[22]. Gatti, L., Guerini, M., & Turchi, M. (2016). SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, *7*(4), 409-421.
[23]. Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings, *30*(1), 25-36. Technical report C-1, the center for research in psychophysiology, University of Florida.
[24]. Stone, P. J., Dunphy, D. C., Smith, M. S., &Oglivie D. M. (1966). The General Enquirer: A computer Approach to Content Analysis. MIT Press, Cambridge MA.
[25]. Tong, R. M. (2001). An operational system for detecting and tracking opinions in on-line discussion. In Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification (Vol. *1*, No. 6).
[26]. Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics (pp. 174-181). Association for Computational Linguistics.
[27]. Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics.
[28]. Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS), 21*(4), 315-346.
[29]. Saif, H., Fernandez, M., He, Y., & Alani, H. (2014). Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In European Semantic Web Conference(pp. 83-98). Springer, Cham.
[30]. Khan, F. H., Qamar, U., & Bashir, S. (2016). SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection. *Applied Soft Computing, 39*, 140-153.
[31]. Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, *1*, 142-150. Association for Computational Linguistics.
[32]. Sharma, S. S., & Dutta, G. (2018). Polarity Determination of Movie Reviews: A Systematic Literature Review. *International Journal of Innovative Knowledge Concepts, 6*(12), 43-55.
[33]. Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management, 50*(1), 104-112.
[34]. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.
[35]. Mullen, T., & Collier, N. (2004). Sentiment Analysis using Support Vector Machines with Diverse Information Sources. In EMNLP, 4, 412-418.
[36]. Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. *In Proceedings of the 14th ACM international conference on Information and knowledge management* (pp. 617-624). ACM.
[37]. Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational linguistics*, *37*(1), 9-27.
[38]. Wiebe, J. (2000). Learning subjective adjectives from corpora. Aaai/iaai, *20*, 1-6.
[39]. Kim, M. S., Kim, J. W., & Jing, C. (2016). Comparison of domain-specific lexicon construction methods for sentiment analysis. *Advanced Science and Technology Letters, 135*, 152-156.
[40]. Khoo, C. S., & Johnkhan, S. B. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science, 44*(4), 491-511.
[41]. Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., &Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(1), 1-29.
[42]. Poria, S., Cambria, E., Gelbukh, A., Bisio, F., & Hussain, A. (2015). Sentiment data flow analysis by means of dynamic linguistic patterns. *IEEE Computational Intelligence Magazine*, *10*(4), 26-36.
[43]. Asghar, M. Z., Khan, A., Ahmad, S., Qasim, M., & Khan, I. A. (2017). Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PloS one, *12*(2).
[44]. Jiménez-Zafra, S. M., Martin, M., González, M. D. M., & Lopez, L. A. U. (2016). Domain adaptation of polarity lexicon combining term frequency and bootstrapping. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 137-146).

[45]. Khan, F. H., Qamar, U., & Bashir, S. (2016). Senti?CS: Building a lexical resource for sentiment analysis using subjective feature selection and normalized Chi-Square-based feature weight generation. *Expert Systems*, *33*(5), 489-500.
[46]. Almatarneh, S., & Gamallo, P. (2017). Automatic construction of domain-specific sentiment lexicons for polarity classification. In International Conference on Practical Applications of Agents and Multi-Agent Systems (pp. 175-182). Springer, Cham.
[47]. Lee, H., Han, Y., Kim, K., & Kim, K. (2014). Sentiment analysis on online social network using probability Model. *In Proceedings of the Sixth International Conference on Advances in Future Internet* (pp. 14-19).
[48]. Labille, K., Gauch, S., & Alfarhood, S. (2017). Creating domain-specific sentiment lexicons via text mining. *In Proc. Workshop Issues Sentiment Discovery Opinion Mining (WISDOM).*
[49]. Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.
[50]. Park, S., Lee, W., & Moon, I. C. (2015). Efficient extraction of domain specific sentiment lexicon with active learning. *Pattern Recognition Letters*, *56*, 38-44.
[51]. Whitelaw, C., Garg, N., &Argamon, S. (2005). Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 625-631). ACM.
[52]. Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems (TOIS), 26*(3), (Page No.)
[53]. Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences, 181(6), 1138-1152.
[54]. Hung, C., & Lin, H. K. (2013). Using objective words in SentiWordNet to improve word-of-mouth sentiment classification. *IEEE Intelligent Systems*, (2), 47-54.
[55]. Moraes, R., Valiati, J. F., &Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications, 40*(2), 621-633.
[56]. Mukherjee, S., & Joshi, S. (2014). Author-Specific Sentiment Aggregation for Polarity Prediction of Reviews. In LREC (pp. 3092-3099).
[57]. Lin, C., He, Y., & Everson, R. (2010). A comparative study of Bayesian models for unsupervised sentiment detection. In Proceedings of the fourteenth conference on computational natural language learning (pp. 144-152). Association for Computational Linguistics.
[58]. He, Y., & Zhou, D. (2011). Self-training from labeled features for sentiment analysis. *Information Processing & Management, 47*(4), 606-616.
[59]. Socher, R., Pennington, J., Huang, E. H., Ng, A. Y., & Manning, C. D. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. In Proceedings of the conference on empirical methods in natural language processing (pp. 151-161). Association for Computational Linguistics.
[60]. Sharma, A., & Dey, S. (2012). Performance investigation of feature selection methods and sentiment lexicons for sentiment analysis. *IJCA Special Issue on Advanced Computing and Communication Technologies for HPC Applications*, *3*, 15-20.
[61]. Singh, V. K., Piryani, R., Uddin, A., & Waila, P. (2013). Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches. In Knowledge and Smart Technology (KST), 5th International Conference on (pp. 122-127). IEEE.
[62]. Dhande, L. L., & Patnaik, G. K. (2014). Analyzing sentiment of movie review data using Naive Bayes neural classifier. *International Journal of Emerging Trends & Technology in Computer Science 3*(4), 313-320.
[63]. Tomar, D. S., & Sharma, P. (2016). A text polarity analysis using SentiWordNet based an algorithm. *International Journal of Computer Science and Information Technologies*, *7*(1), 190-193.
[64]. Porter, M. F. (2001). Snowball: A language for stemming algorithms, 1-20.
[65]. Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, 1, 173-180). Association for Computational Linguistics.
[66]. Sahu, Y., Thakur, G.S., & Dhyani S. (2019). Dynamic feature based computational model of sentiment analysis to improve teaching learning system. *International Journal on Emerging Technologies*, *10*(4), 17-23.

How to cite this article: Sharma, Shashank and Dutta, Gautam (2020). A Novel Approach for Creating Sentiment Lexicon for Movie Reviews using Star Ratings. *International Journal on Emerging Technologies*, *11*(1): 376–386.