Efficient Feature Extraction in Sentiment Classification for Contrastive Sentences

Sonu Lal Gupta
Gautam Buddha University, Greater Noida, India-201308
Email: sonugupta2006@gmail.com

Anurag Singh Baghel
Gautam Buddha University, Greater Noida, India-201308
Email: anuragsbaghel@gmail.com

Abstract—Sentiment Classification is a special task of Sentiments Analysis in which a text document is assigned into some category like positive, negative, and neutral on the basis of some subjective information contained in documents. This subjective information called as sentiment features are highly responsible for efficient sentiment classification. Thus, Feature extraction is essentially an important task for sentiment classification at any level. This study explores most relevant and crucial features for sentiment classification and groups them into seven categories, named as, Basic features, Seed word features, TF-IDF, Punctuation based features, Sentence based features, N-grams, and POS lexicons. This paper proposes two new sentence based features which are helpful in assigning the overall sentiment of contrastive sentences and on the basis of proposed features; two algorithms are developed to find the sentiment of contrastive sentences. The dataset of TripAdvisor is used to evaluate our proposed features. Obtained results are compared with several state-of-the-art studies using various features on the same dataset and achieve superior performance.

Index Terms—Sentiment analysis, Sentiment classification, Contrastive sentences, Review subjectivity, Polarity detection, Machine learning, Lexicon.

I. INTRODUCTION

The Web is a pool of online information which consists of text data i.e. facts and reviews or opinions about those facts. Facts are objective sentences which are based on proof and do not have any sentiments while opinions are subjective sentences which brief about different sentiments of different people towards entities. Processing the opinions is commonly known as sentiments analysis which has attained a high popularity in the last decade because of the rise in social media. It aims to determine the attitude of a person on the web in terms of some topics or overall opinion for a document.

Sentiment classification is such a task which labels various documents into categories like positive sentiments, negative sentiments or neutral as per opinion information contains in the documents [1-2]. Sentiment classification may be broadly categorized into levels namely document level, sentence level, and finally aspect/feature level [3-6]. Document level specifies the document polarity as positive or negative considering the document as a single unit, while sentence level considers the whole sentence and expresses the sentiment. The aspect level analysis first identifies the entities and further opinions about those entities.

Every text consists of certain features which express the sentiment of the text. Features may express sentiments implicitly or explicitly. For classification of sentiments, a feature is nothing but a piece of sensible information from the text which could be a word or a combination of many words or a full sentence which brings up the definition of the polarity of the text in terms of the positive, negative or neutral review. Feature extraction is a necessary step in sentiment classification to extract the most representative features which are helpful in distinguishing classes [7]. Almost every text contains enormous features out of which around seventy percent features are irrelevant and creates noise. The main purpose of feature extraction is to find as many relevant features which could speed up the process of classification of data. The more accurate is the extraction of features the more accurate will be the sentiment analysis.

Sentiments are not always expressed explicitly in sentences. Like in sentence "how can anyone purchase this item?" sentiment is negative but implicit and has no words which contains sentiments directly. The polarity of many words highly depends on the domain. Their polarity cannot be fixed and changes from domain to domain. Presence of sarcasm and negative sentences are also a big threat for accurate sentiment classification. Similarly, contrastive sentences are the major challenge to find the overall sentiment of the sentence. In this research, we are proposing two novel algorithms to assign the overall sentiment of contrastive sentences.

Sentiment analysis techniques are broadly categorized into three techniques: machine learning based, lexicons

dictionary and hybrid based. In this paper, hybrid based approach is used to detect the polarity of sentences as using machine learning based and lexicon dictionary-based approaches alone has their own demerits. Support vector machine (SVM) and linear regression are the being used as machine learning classifier and an inbuilt dictionary SentiWordNet [8] is used to obtain the semantic angle of words in a review by obtaining word polarities which is the standard dictionary available today.

This paper makes the following contributions.

1. It studies and analyzes the various sentiment features and categorizes them into seven categories.
2. It proposes two new sentences based features to find out the sentiments of contrasting sentences and develops the algorithm.
3. It performs the experiment on TripAdvisor dataset to evaluate the proposed features using SVM and linear regression classifiers.
4. It compares the obtained results with previous experimented results on TripAdvisor datasets.

Rest of the paper is organized as follows. Section II reviews the previous studies on sentiment analysis using different features. Various features are explored and grouped in section III. Our proposed features and algorithms are explained in section IV and evaluated and compared with other results in section V. Finally, section VI concludes the research work and suggests some future work.

II. RELATED WORK

A lot of studies have been done to find out the sentiments of documents at all levels. Experiments have been conducted especially on Chinese text reviews [9-10], movie reviews [11], hotel reviews [12], product reviews [13-14], and Twitter data [15-17]. Feature extraction is also explored in image datasets [18].

Gindl et al. [19] presented a technique which recognizes flimsy contextualization and refines the contextualized supposition word references and expels the destructive terms from the contextualized notion dictionary, making a bland space free vocabulary. The assessments show that such a preprocessing of contextualized estimation dictionaries essentially enhances the execution of a sentiment determination technique.

Bespalov et al. [20] propose a proficient implanting for demonstrating higher request (n-gram) phrases for the problem of sentiment classification that ventures the n-grams to low-dimensional inert semantic space, where a function for classification can be defined. They use a deep neural network to assemble a bound together discriminative structure that takes into account evaluating the parameters of the latent space and classification function. They assess the execution of the proposed strategy on two benchmark datasets of Amazon and TripAdvisor.

A domain-specific corpus and lexicon-based approach for sentiment classification of reviews were used by Grabner et al. [21]. In their work, they prepare a corpus from the specific dataset by using semantic orientation and classified the customer reviews and obtained results were compared with other findings.

Gezici et al. [22] propose and explore new features to be utilized as a part of a word review polarity based approach to deal with sentiment determination. Initially, all the sentences were extracted from the document for sentence level classification and the overall sentiment of the review was evaluated. They utilized distinctive parts of the sentence, for example, length, immaculateness, irrealis substance, and position inside the stubborn content to discover sentences which are helpful in finding better sentiment overall. The experiment was performed on TripAdvisor dataset to evaluate the significance of sentence-level features on sentiment classification.

Agarwal et al. [23] extracted various features from text like dependency features, unigrams or bi-grams. This work also highlighted bi-tagged features which are based on predetermined POS patterns and multiple composite features were formed. Two feature selection methods, Information gain (IG) and minimum redundancy maximum relevancy (mRMR), are used to eliminate the noisy and irrelevant features from the feature vector. Machine-learning classifiers like Multinomial Naive Bayes (BMNB) and bolster vector machine (SVM) were utilized for classification of the text documents into various classes. Experiment on various categories of features was performed on movie review datasets and product review datasets like book, DVD, and electronics.

An algorithm for sentiment analysis of Chinese product reviews was proposed by Lizhen et al. [9]. In this algorithm, they consider a vector model based on features extraction. In their proposal relationships between words were identified in various ways which were represented by either adverb of degree, punctuations or over–modifiers. The proposed model was evaluated on a dataset of 3500 documents from the Chinese corpus.

Sharma et al. [11] proposed a system which classifies the polarity of the movie reviews on the basis of features by handling negation, intensifier, conjunction and synonyms with appropriate pre-processing steps. They have used SentiWordNet tool for calculating the scores of reviews.

A classification approach by Ding et al. [24] was used to solve two tasks of sentiment analysis: identifying opinion sentence and judging sentiment polarity of the emotional sentence. They have used Jieba techniques to preprocess the micro-blog texts then extracts features which include sentiment lexicons based features. They employed seven classifiers (SVM with linear kernel, SVM with polynomial kernel, SVM with RBF kernel, K Nearest Neighbor, Decision Tree, Naïve Bayes and Random Forest) to train the classification models respectively and compares their experimental results. The experimental result shows that Random Forest classifier achieves the best performance.

III. Features for Sentiment Classification

In the journey of finding accurate sentiment classification, many researchers have explored and proposed various types of features. Initially, Pang et al. [25] used unigram, bi-gram, and adjectives based features for the task of sentiment analysis. As per the need of the hour, various domain-specific features like punctuations based, Part of Speech (POS) based and seed words were proposed and utilized in the study of many researchers. In this section, after studying and analyzing various research papers and studies, we are summarizing and categorizing various features important for sentiment classification as per Table 1.

Table 1. Various Features for Sentiment Analysis

S.N.	Categories	Feature	Feature Description
1	Basic Features	F1	Average review polarity
		F2	Review Purity
		F3	Review subjectivity
2	Seed words Features	F4	Frequency of seed words
		F5	Average sentiment of seed words
		F6	Standard deviation of sentiment of seed words
3	TF-IDF	F7	Total TF-IDF scores of all words
		F8	Average sentiment of reviews weighted by scores of TF-IDF
4	Punctuation based features	F9	Number of all exclamation marks
		F10	Number of all question marks
		F11	Number of all positive smiley's
		F12	Number of negative smiley's
5	Sentence based features	F13	Average polarity of first line
		F14	Average polarity of last line
		F15	First line review purity
		F16	Last line review purity
		F17	Total scores of TF-IDF of words coming in the first line
		F18	TF-IDF polarity weighted by scores of first line
		F19	Total scores of TF-IDF of words coming in the last line
		F20	TF-IDF polarity weighted by scores of last line
		F21	Total number of all sentences in the review
		F22	Average review polarity of subjective sentences
		F23	Average review polarity of pure sentences
		F24	Average polarity of realistic sentences
6	N-Gram features	F25	Unigram, Bigram, Trigram
7	Part-of-Speech (POS)	F26	Adjective
		F27	Adverb
		F28	Verb
		F29	Noun

For classification of sentiments in a given document or a review, we have explored twenty-nine features which are grouped under seven main categories as Basic features, Seed word features, TF-IDF, Punctuation based features, Sentence based features, N-grams and, Part-of-Speech lexicons.

Before going into details of these features, A text document D is defined as a collection of sentences $D = L_1L_2L_3...L_M$, where M is the total number of sentences in document D. Similarly, each sentence L_i is a collection of ordered words $L_i = w_{i1}w_{i2}...w_{iN(i)}$ where N(i) is the total number of words present in sentence L_i. The document D can also be represented as a collection of ordered words $w_1w_2w_3...w_T$, where T is the total number of words in the text document D [22].

A. Basic Features

In this group of features, we exploit review polarity, review purity, and review subjectivity of the text which are the most common and straightforward features in text classification and are being used several times in the literature. In the given formula $pol(w_j)$ denotes the dominant polarity of word w_j of D as obtained from SentiWordNet, and $|pol(w_j)|$ denotes the absolute polarity of word w_j.

Table 2. Basic Features

F1	Average polarity of reviews	$\frac{1}{T}\sum_{J=1...T} pol(wj))$
F2	Purity of reviews	$\sum_{J=1...T} pol(wj)/\sum_{J=1...T}|pol(wj)|)$
F3	Review subjectivity	1 if review is subjective

i. Feature F1 is based on the concept of average review polarity (AP). A word w is decided as positive if Pol(w) > 0, and decided as negative if Pol(w) < 0.

ii. Feature F2 is review purity which is the ratio of absolute polarity and dominant polarity and

iii. Feature F3 is Review subjectivity which is one, a binary variable if any one of the sentences in a review is subjective.

B. Seed Words

In this category of features, we have set of two seeds one as positive seed words and other as negative. These words which are called as seed words are globally accepted as positive or negative words irrespective of the content of the sentence.

These words can be defined easily on the manual basis or using supervised learning approach. Several research works are done to create a seed list as a thesaurus or lexical database e.g. WordNet [26] as a seed list of words [10]. Another famous seed word list is proposed in [27].

These seed words features ultimately help in performing calculations based on the occurrences of these

words in a review to gain some clues for determination of sentiment. Here we have listed a small set of seed words, 20 positives and 20 negatives, to depict the seed word features F4, F5, and F6 in Table 3.

Table 3. Positive and Negative Seed Words

Positive Words	Negative words
Great	Never
Excellent	Worst
Wonderful	Bad
Perfect	Even
Fantastic	Terrible
Wonderful	Rude
Comfortable	Poor
Helpful	Disappointment
Friendly	Upset
Lovely	Filthy
Glad	Vicious
Admirable	Corrupt
Amaze	Inferior
Appeal	Rotten
Bliss	Foul
Bright	Boring
Powerful	Stressed
Secure	Weird
Romantic	Criticism
Upright	Disgusted

We define seed W(R) as the set of seed words that appear in review R and extract three features which are related to seed words in the review of the text. Feature F4 is the frequency of appearance of positive seed words and negative for every opinion coming in the data set. Another seed word feature F5 is the average polarity of seed words and standard deviation of the review polarity of the seed words is feature F6.

Table 4. Seed Word Features

F4	Frequency of Seed words	$|Seed\ W(R)|/|R|$
F5	Average Polarity of Seed words	$\dfrac{1}{|seed\ W(R)|}\displaystyle\sum_{wj\in Seed\ W(R)} pol(wj)$
F6	Std dev of Polarity of Seed words	σ_i $= (((((\mu_i - Positive\_Score(t_i)^2$ $* Positive\_Count(t_i))$ $- (\mu_i - NegativeScore(t_i))^2$ $* Negative\_Count(t_i)))$ $/Positive\_Count(t_i)$ $+ Negative\_Count(t_i)))^{1/2}$

C. TF-IDF

In this group, we consider features, total TF-IDF scores of all words (feature F7) and average review polarity weighted by delta TF-IDF scores (feature F8). TF-IDF score of a word-sense pair is computed of features based

on the relative occurrence of a word-sense among positive and negative classes. Term frequency (TF) weights indicate the relative importance of features in document representations.

Table 5. TF-IDF Based Features

F7	Total tf-idf scores of all words	Δ tf *idf(wi) is defined as tf Δ idf(wi) = tf Δ idf(wi,+)−tf Δidf(wi,−)
F8	Average review polarity weighted by tf-idf scores	$\dfrac{1}{|subjW(R)|}\displaystyle\sum_{wj\in subjw(R)} pol(wj)$

D. Punctuation-based Features

This group of features is based on punctuation present in the text. These features consider the number of question in the message and the number of exclamation marks in the message. These features are very useful especially for Twitter as they may give some information about the sentiment of a review. We have also explored two new features number of positive smileys and number of negative smileys.

Table 6. Punctuations Based Features

F9	Number of exclamation marks
F10	Number of question marks
F11	Number of positive smileys
F12	Number of negative smileys

E. Sentence-based Features

In this group of features, features are extracted based on sentence type e.g. subjective, pure, and realistic and sentence position e.g. first line and last line. Features include several basic ones such as the average polarity of the first sentence and the average polarity of subjective or pure sentences. Also, compute TF-IDF scores of words in first and last line on sentence based features. Table 6 shows all 12 Sentence level based features.

F. N-gram features

In this group, features are divided into unigram, bigram, and trigram and so on. N-grams are basically a set of co-occurring words within a given text and when computing the n-grams we typically move one word forward. Unigrams, bigrams, and trigrams of a review are being used to assign the score to a review and thus classify it as positive or negative. Text categorization on the basis of n-gram approach is one of the fastest and robust methods. These features determine the sentiment of review in most simple manner and also work with data that consist of errors and noise like email and newsgroups or blogs.

G. Parts of Speech

Table 7. Punctuations Based Features

F13	Average first line polarity	$\frac{1}{|S_1|}\sum_{w \in S_1} pol(w)$
F14	Average last line polarity	$\frac{1}{|S_1|}\sum_{w \in S_M} pol(w)$
F15	First line purity	$[\sum_{w \in S_1} pol(w)]/[\sum_{w \in S_1}|pol(w)|]$
F16	Last line purity	$[\sum_{w \in S_M} pol(w)]/[\sum_{w \in S_M}|pol(w)|]$
F17	Total tf-idf scores of words in the first line	$\sum_{w \in S_1} \Delta t_f * idf(w) X\, pol(w)$
F18	tf-idf weighted polarity of first line	$\sum_{w \in S_1} \Delta t_f * idf(w)$
F19	Total tf-idf scores of words in the last line	$\sum_{w \in S_M} \Delta t_f * idf(w) X\, pol(w)$
F20	Tf-idf weighted polarity of last line	$\sum_{w \in S_M} \Delta t_f * idf(w)$
F21	Number of sentences in review	M
F22	Average polarity of subjective sentences	$\frac{1}{subjS(R)}\sum_{w \in subjW(R)} pol(w)$
F23	Average polarity of pure sentences	$\frac{1}{pureS(R)}\sum_{w \in pure(R)} pol(w)$
F24	Average polarity of realistic sentences	$\frac{1}{nonIrS(R)}\sum_{w \in nonIr(R)} pol(w)$

In this group of features, we have exploited four set of features: Adjective (F26), Adverb (F27), Verb (F28), Noun (F29). Parts of Speech (POS) information is very common in natural language processing tasks. One of the most important reasons is that they provide a very simple and rudimentary form to avoid ambiguity in the sentiment of a word.

Adjectives are the most frequently used features of all POS features. Researchers show that even only adjectives in a sentence depict very high accuracy for feature generation resulting useful in text classification. Pang Lee et al. [25] achieved an accuracy of around 82.8% in movie review domains using only adjectives in movie review domains. Further, Turney [28] worked on POS information. He used some tag patterns with a window of maximum three words that are till trigrams. In his experiments, he considered JJ(adjective), RB(adverb), NN(single common nouns), NNS(plural common nouns) POS-tags with some set of rules for classification.

Only adverbs show no prior polarity in feature generation. But when they occur with sentiment-bearing adjectives, they can play a major role in determining the sentiment of a sentence. Benamara et al. [29] have shown how the adverbs alter the sentiment value of the adjective that they are used with.

Research work also shows that other than the adjective, adverb, verb, and noun also plays as an important feature. But if words that are adjectives, adverbs, and nouns are taken in combination then results are best and secondly, with verbs, shows that these two parts of speech are indeed more helpful in polarity classification.

IV. PROPOSED FEATURES

In this section, proposed features are explained to evaluate the sentiment of contrastive sentences. This work focuses on the problem of detecting sentiments of contrastive sentences in a given review based on proposed approach. We analyzed that sentences with few contrastive conjunctions like "but", "however", and "although/though", connect ideas that contrast. It can altogether change the meaning of the sentence changing its polarity too. For example "*this movie was excellent but it was lengthy and serious*". Here the movie rating is suddenly dropped as it has many negative points along with positive features.

We consider the problem of contrastive sentences annotating the polarity of the sentence. Sentence level sentiment classification approaches, whether machine learning based or lexicon based, are unable to consider the word structure of contrastive sentences and results into misclassification hence poor classification performance. By motivating from the fact, we propose two new features which take the word structure of contrastive sentences into account. The proposed features slice the sentence at the place where conjunction occurs and use words before and after contrastive word as features. Proposed features are names as F30 and F31.

A. Proposed Feature F30

The proposed feature is inspired by the human behavior in expressing their opinions. As humans express their opinions about any product or service which may be positive or negative or mixed in nature. When an opinion is mixed then the human tendency is to express the most influential emotions at the end.

As contrastive sentences have two phrases: phrase1 before the conjunction word and phrase2 after the conjunction word. In this feature, if a conjunction word occurs in a sentence, the sentence is split into two phrases and phrase1 with conjunction word is discarded from the sentence. The only phrase2 is retained for further sentiment classification. It has the implication that the sentiment of the only phrase2 will be the overall all sentiment of the whole sentence. This feature shortens the length of sentence thus increases the speed and accuracy of the classification process. The machine learning approach is used to evaluate the proposed feature F30.

B. Proposed Feature F31

The proposed feature F31 is based on the lexicon approach of sentiment classification where SentiWordNet is being utilized as lexicon resource. In this feature, the contrastive sentence is also split into two phrases like in feature F30. In phrase 1, the positive and negative scores of all the lexicons according to their POS-tag are extracted from SentiWordNet and scores are summed up. If more than one synsets are found in lexicon resource then weighted average score is considered. The difference of summed scores of all the lexicons is the sentiment of phrase1. The similar procedure is repeated for phrase 2 also. The absolute score of phrase 1 and phrase 2 is compared and the sentiment of the phrase with the higher score is the overall sentiment or polarity of the contrastive sentence.

Algorithm 1. The Polarity of the Contrastive Sentences

Input: Contrastive sentence S with conjunction c and lexicon resource

Output: Sentence Polarity as positive or negative

Step 1: Take phrase 1 as P1and Phrase 2 as P2

Step 2: Find the sum of positive scores of all lexicons of P1according to POS-tag i.e. $\sum Ps$

Step 3: Find sum of negative scores of all lexicons according to POS-tag i.e. $\sum Ns$

Step 4: IF more than one synsets are found in lexicon resource THEN

Step 5: Consider weighted average score. i.e Avg(ws)

Step 6: Compute the difference of summed scores SP1 = $\sum Ps$ -$\sum Ns$

Step 7: IF SP1> 0 then sentiment of P1 is positive

ELSE   sentiment of P1 is negative

Step 8: Repeat the steps 2 to 7 for P2

Step 9: IF abscore(SP1) > abscore(SP2) THEN

      Overall sentiment of S is as of P1

ELSE   Overall sentiment of S is as of P2

END IF

The advantages of using this feature are that it considers both the phrases and provides the weight to the number of lexicons appeared before and after the conjunction word. The approach has been summarized in Algorithm 1.

V. Experiment and Results

In this section, we intend to evaluate the proposed features and other features studied above for the sentiment classification. In the provided experiment we have taken the data set of hotel reviews from TripAdvisor.

A. Dataset

TripAdvisor is one of the most popular travel social network websites. TripAdvisor contains millions of written and ranked reviews about restaurants, hotels, and attractions from a large number of travelers over the world. Tourists are able to plan their trip checking information, ranking list and experiences from others. In this website, users write opinions of 100 character minimum and rank them with 1 to 5 score (1 is representing a terrible assessment and 5 an excellent assessment). TripAdvisor has, therefore, become a rich source of data for SA research and applications [30].

Dataset used in this study consists of 10000 text reviews. Each review is rated from one star (Terrible) to five star (Excellent). We have chosen 5000 positive reviews and 5000 negative reviews for our experiment.

B. Experimental set-up

We have performed many experiments to evaluate the impact of proposed features on the considered dataset. To evaluate the proposed feature F30, a supervised machine learning approach is used in which all the sentences were extracted from each review and each sentence was preprocessed and passed to supervised machine learning classifier for classification. Then sentiment of each sentence was aggregated to determine the overall sentiment of the text document. The proposed feature F31 is a lexicon based feature in which the contrastive sentences are classified on the basis of lexicon-based approach and rest of the sentences are being classified by underlying machine learning approach so in this way, a hybrid classification approach is suggested to evaluate the proposed feature F31. First, all contrastive sentences are segregated, if they are any, in a text document and classified as approach explained in algorithm 1 while remaining sentences are classified by underlying machine learning classifier. The overall sentiment of the text document is determined by aggregating the sentiments of all sentences.

We have considered two machine learning classifiers SVM and linear regression for the classification task. The reason behind selecting these classifiers is that SVM is considered best to handle large feature spaces and limiting over-fitting simultaneously, while Logistic Regression is a simple, and commonly used, well-performing classifier. We have tested our proposed features along with many features among 29 features discussed in this study.

Table 8. Experimental Results on TripAdvisor Dataset

Features used in experiment	SVM			Linear Regression		
	Precision	Recall	F-measure	Precision	Recall	F-measure
TF-IDF + unigram without POS	86.72%	82.64%	84.63%	85.12%	78.62%	81.74%
TF-IDF + unigram with POS	89.34%	85.65%	87.45%	87.86%	81.76%	84.70%
TF-IDF + unigram with POS + Proposed feature F30	92.45%	88.85%	90.61%	89.63%	84.17%	86.81%
TF-IDF + unigram with POS + Proposed feature F31	94.24%	91.58%	92.89%	93.32%	89.93%	91.59%

The measure of performance was chosen as the accuracy of classification (F-measure), precision and recall in our experiment given in (1)-(3).

$$precision = \frac{tp}{tp+fp} \qquad (1)$$

$$recall = \frac{tp}{tp+fn} \qquad (2)$$

$$F - measure = \frac{2*precision*recall}{precision+recall} \qquad (3)$$

The experimental results obtained are displayed in Table 8. From the above Table 8, the following conclusions can be drawn.

(a) The feature TF-IDF achieves higher classification accuracy along with POS rather using alone in both the classifiers. The accuracy is improved by 2.82% in SVM and 2.96% accuracy is increased in linear regression using POS with TF-IDF.
(b) After incorporating our proposed feature F30 with TF-IDF and POS, a significant improvement was observed in precision, recall, and accuracy in both the classifiers. The accuracy of 3.16% was increased in SVM using proposed feature F30 along with TF-IDF and POS and The accuracy of 2.11% was increased in case of linear regression using proposed feature F30 along with TF-IDF and POS.
(c) The proposed feature F31gives the best results when used with TF-IDF and POS. The accuracy of 5.44% was increased in SVM and the accuracy of 6.89% was increased in linear regression using proposed feature F31 along with TF-IDF and POS.
(d) It is recommended to use the POS feature rather using all the lexicons along with other features as POS improves the performance of the classifier in terms of accuracy by discarding the words which don't possess any sentiments. This also reduces the size of datasets so time complexity is also improved.
(e) SVM classifier is the most performing one in our case as SVM is providing not only higher classification accuracy but also higher precision and recall with all the considered features.

Furthermore, we have compared our obtained results with the results of previous researches conducted on TripAdvisor datasets and have summarized in Table 9.

Table 9. Comparative Performance of Sentiment Classification System on TripAdvisor Dataset

Previous work	F-measure	Error rate
[22]	81.45%	-
[31]	82%	-
[20]	-	7.37%
[21]	69%	-
[19]	79%	-
The proposed work	92.89%	-

From the Table 9, this can be easily observed that our proposed features improve the classification accuracy significantly and achieve highest classification accuracy so far.

VI. CONCLUSION AND FUTURE WORK

In the era of information fusion, users are writing reviews for hotels, products, movies and other services more and more. Sentiment classification is to classify these reviews into some categories positive, negative or neutral based on the sentiment features contained in reviews. The more powerful feature extraction will lead us to more accurate sentiment classification. Various features have been studied and explored for efficient sentiment classification. We have grouped them into seven categories, named as, Basic features, Seed word features, TF-IDF, Punctuation based features, Sentence based features, N-grams, and POS lexicons. Our study makes a contribution towards efficient sentiment classification by proposing two novel sentence based features and developing algorithms for finding sentiments. Proposed features are capable to classify the contrastive sentences which involve contrastive conjunctions.

To evaluate our proposed features, we have classified the TripAdvisor dataset along with other features using two machine learning classifiers SVM and linear regression. Obtained results are compared with five state-of-the-art results on TripAdvisor datasets and demonstrate the superiority of proposed features in both the classifiers. In this experiment F-measure accuracy is

considered as a measure of performance and SVM achieves the accuracy of 92.89% and linear regression achieves the accuracy of 91.59% which is the best performance so far.

In future, we intend to extract more domain specific features and study the impact of machine learning approaches on sentiment classification.

REFERENCES

[1] B. Narendra et al., "Sentiment Analysis on Movie Reviews: A Comparative Study of Machine Learning Algorithms and Open Source Technologies," International Journal of Intelligent Systems and Applications, vol. 8, pp. 66-70, 2016.

[2] Bing Liu, "Sentiment Analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1-167, 2012.

[3] Subhabrata Mukherjee and Pushpak Bhattacharyya, "Sentiment analysis: A literature survey," arXiv preprint arXiv:1304.4520, 2013.

[4] Akshi Kumar and Teeja Mary Sebastian, "Sentiment analysis: A perspective on its past, present and future," International Journal of Intelligent Systems and Applications, vol. 4, no. 10, pp. 1-14, 2012.

[5] Samina Khalid, Tehmina Khalil, and Shamila Nasreen, "A Survey of Feature Selection and Feature Extraction techniques in machine learning," in In Science and Information Conference (SAI), 2014, pp. 372-378.

[6] Samina Khalid, Tehmina Khalil, and Shamila Nasreen, "A review on feature extraction and feature selection for handwritten character recognition," International Journal of Advanced Computer Science and Applications, vol. 6, no. 2, pp. 204-215, 2015.

[7] Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, and Fazal Masud Kundi, "A Review of Feature Extraction in Sentiment Analysis," Journal of Basic and Applied Scientific Research, vol. 4, no. 3, pp. 181-186, 2014.

[8] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," In LREC, vol. 10, pp. 2200-2204, 2010.

[9] Liu Lizhen, Song Wei, Wang Hanshi, Li Chuchu, and Lu Jingli, "A Novel Feature-based Method for Sentiment Analysis of Chinese product reviews," China communications, vol. 11, no. 3, pp. 154-164, 2014.

[10] Xiaowen Ding, Bing Liu, and Philip S. Yu, "A holistic lexicon-based approach to opinion mining," in WSDM '08 Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008, pp. 231-240.

[11] Pallavi Sharma and Nidhi Mishra, "Feature level Sentiment Analysis on Movie Reviews," in 2nd International Conference on Next Generation Computing Technologies (NGCT-2016), 2016, pp. 306-311.

[12] Ana Valdivia, M. Victoria Luzón, and Francisco Herrera, "Sentiment Analysis on TripAdvisor: Are There Inconsistencies in User Reviews?," in International Conference on Hybrid Artificial Intelligence Systems. HAIS, vol. 10334, 2017, pp. 15-25.

[13] Yanyan Meng. (2012) Sentiment analysis: A study on product features. Dissertations, Theses, and Student Research from the College of Business. 28.

[14] Yelena Mejova and Padmini Srinivasan, "Exploring Feature Definition and Selection for Sentiment Classifiers," in Fifth International AAAI Conference on Weblogs and Social Media, 2011, pp. 546-549.

[15] Gizem Gezici, Rahim Dehkharghani, Berrin Yanikoglu, Dilek Tapucu, and Yucel Saygin, "SU-Sentilab : A Classification System for Sentiment Analysis in Twitter," in In SemEval@ NAACL-HLT, 2013, pp. 471-477.

[16] Munir Ahmad and Shabib Aftab, "Analyzing the Performance of SVM for Polarity Detection with Different Datasets," International Journal of Modern Education and Computer Science, vol. 9, no. 10, pp. 29-36, 2017.

[17] Akhilesh Kumar Singh, Deepak Kumar Gupta, and Raj Mohan Singh, "Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly Election, 2017," International Journal of Modern Education and Computer Science, vol. 9, no. 9, pp. 60-68, 2017.

[18] A Setiyoko, I G W S Dharma, and T Haryanto, "Recent development of feature extraction and classification multispectral/hyperspectral images: a systematic literature review," In Journal of Physics: Conference Series, vol. 801, no. 1, 2017.

[19] Stefan Gindl, Albert Weichselbraun, and Arno Scharl, "Cross-Domain Contextualization of Sentiment Lexicons," in Proceeding of the 2010 conference on ECAI 2010: 19th European Conference on Artificial Intelligence, 2010, pp. 771-776.

[20] Dmitriy Bespalov, Bing Bai, Yanjun Qi, and Ali Shokoufandeh, "Sentiment classification based on supervised latent n-gram analysis," in CIKM '11 Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 375-382.

[21] Dietmar Gräbnera, Markus Zanker, Günther Fliedl, and Matthias Fuchs, "Classification of Customer Reviews based on Sentiment Analysis," in 19th Conference on Information and Communication Technologies in Tourism (ENTER), 2012, pp. 460-470.

[22] Gizem Gezici, Berrin Yanikoglu, Dilek Tapuc, and Yucel Saygın, "New Features for Sentiment Analysis: Do sentences matter?," in CEUR Workshop Proceedings, 2012, pp. 5-15.

[23] Basant Agarwal and Namita Mittal, "Prominent feature extraction for review analysis: an empirical study," Journal of Experimental & Theoretical Artificial Intelligence, vol. 28, no. 3, pp. 485-498, 2014.

[24] Jiayuan Ding, Yongquan Dong, Tongfei Gao, Zichen Zhang, and Yali Liu, "Sentiment Analysis of Chinese Micro-blog based on Classification and Rich Features," in Web Information Systems and Applications Conference, vol. 13th, 2016, pp. 61-66.

[25] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79-86.

[26] George A. Miller, "WordNet: A Lexical Database for English," Communications of the ACM, vol. Vol. 38, No. 11, pp. 39-41, 1995.

[27] Minqing Hu and Bing Liu, "Mining and summarizing customer reviews," in In Proceedings of the 10th ACM SIGKKD International Conference on Knowledge Discovery and Data Mining, Seattle,Washington, USA, 2004, pp. 168-177.

[28] Peter D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in In Proceedings of the 40th annual meeting on association for computational linguistics, 2002, pp. 417-424.

[29] Farah Benamara, Carmine Cesarano, Antonio Picariello, and Diego Reforgiato, "Sentiment analysis: Adjectives and adverbs are better than adjectives alone," In ICWSM, pp. 1-7, 2007.

[30] Hong-yu Zhang, Pu Ji, Jian-qiang Wang, and Xiao-hong Chen, "A novel decision support model for satisfactory restaurants utilizing social information: A case study of TripAdvisor.com," Tourism Management, vol. 59, pp. 281-297, 2017.

[31] Raymond Yiu Keung Lau, Chun Lam Lai, Peter B. Bruza, and Kam F. Wong, "Leveraging web 2.0 data for scalable semi-supervised learning of domain-specific sentiment lexicons," in CIKM '11 Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 2457-2460.

Authors' Profiles

Sonu Lal Gupta received his M.Sc.(Mathematics) degree from Himachal Pradesh University, Shimla, India, and M.Tech. (Computer Science) degree from Jawaharlal Nehru University, Delhi, India. Currently, he is a research fellow at the school of information and communication technology, Gautam Buddha University, Greater Noida, India.

His research includes Machine Learning, Evolutionary Computation, and Big data.

Anurag Singh Baghel has completed his M.Tech (Electronics) in 2000 and D.Phil in 2010 both from University of Allahabad, Allahabad, India. He served as Lecturer (Electronics) from 2004 to 2011 in Banasthali University, Tonk, India and since then he is working as Assistant Professor (Computer Science) in Gautam Buddha University, Greater Noida, India. His areas of interest are – Metaheuristics and applications, Software Engineering, and Big Data. He has published more than 40 research publications in various journals and international conferences. He has supervised more than 40 M.Tech Dissertations. Presently, seven scholars are pursuing Ph.D. under his supervision.