International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6607
***Article Received**: 19 February 2023 **Revised**: 04 April 2023 **Accepted**: 22 April 2023*

Sentiment Classification Using a Sense Enriched Lexicon-based Approach

Anima Srivastava[1], Amit Srivastava[2], Tanveer J. Siddiqui [3]
Department of Electronics and Communication, University of Allahabad, Prayagraj INDIA
anima24.jk@gmail.com[1],aksrivastava@allduniv.ac.in[2],tjsiddiqui@allduniv.ac.in[3]

Abstract— The prominent approach in sentiment polarity classification is the Lexicon-based approach which relies on a dictionary to assign a score to subjective words. Most of the existing work use score of the most dominant sense in this process instead of using the contextually appropriate sense. The use of Word Sense Disambiguation (WSD) is less investigated in the sentiment classification tasks. This paper investigates the effect of integrating WSD into a Lexicon-based approach for Sentiment Polarity classification and compares it with the existing Lexicon-based approaches and the state-of-art supervised approaches. The lexicon used in this work is SentiWordNet v2.0. The proposed approach, called Sense Enriched Lexicon-based Approach (SELSA), uses a word sense disambiguation module to identify the correct sense of subjective words. Instead of using the score of the most frequent sense, it uses the score of the contextually appropriate sense only. For the purpose of comparison with the supervised approaches, the authors investigate Naïve Bayes (NB) and Support Vector Machines (SVM) classifiers which tend to perform better in earlier research. The performance of these classifiers is evaluated using Word2vec, Hashing Vectorizer, and bi-gram feature. The best-performing classifier-feature combination is used for comparison. All the evaluations are done on the Movie Review dataset. SELSA achieves an accuracy of 96.25% which is significantly better than the accuracy obtained by SentiWordNet-based approach without WSD on the same dataset. The performance of the proposed algorithm is also compared with the best-performing supervised classifier investigated in this work and earlier reported works on the same dataset. The results reveal that the SVM classifier performs better than SentiWordNet approach without WSD. However, after incorporating WSD the performance of the proposed Lexicon-based approach is significantly improved and it surpasses the best-performing supervised classifier (SVM with bi-gram features).

Keywords- Sentiment Analysis, Sentiment Classification, IMDB dataset, SELSA, Machine Learning.

I. INTRODUCTION

Nowadays, millions of people use social networking websites to share their views on products, events, government policies, economic and political reforms, elections, movies, etc. leading to the proliferation of opinionated content on the web. This content is publically accessible and acts as a guiding resource to many people in making a decision, especially related to the purchase. The sentiment of people's views is also used by companies, service providers, policymakers, government, and political parties in making decisions to shape their future. However, this huge amount of content cannot be analyzed manually. This has made automatic sentiment analysis or opinion mining a hot topic of research. Sentiment Analysis (SA) is concerned with the automatic analysis of opinions, views, emotions, and attitudes expressed in the text. It is one of the most heavily researched areas of Natural Language Processing nowadays, owing to the availability of a huge amount of freely available opinionated content and its useful applications.

A lot of work has been done on sentiment analysis. A sizeable amount of this work focuses on sentiment polarity classification, i.e., classifying text as positive or negative based on sentiment polarity. A dominant approach in SA is to rely on some Lexicon to extract sentiment-bearing words and use them in classification. SentiWordNet is one such lexicon that is used by many researchers [1, 2, 3, 4 5]. It associates each synonym set in WordNet with a sentiment score and polarity. Most of the SentiWordNet-based sentiment classification systems simply extract the sentiment score of the first sense of the word listed in SentiWordNet or opt for the average score instead of disambiguating the word and using the score of the contextually appropriate sense. However, different senses of a term may convey sentiments of varying strength. The sense in which a word is being used may affect the subjectivity score and may lead to incorrect classification. For example, the word "active" has multiple senses listed in WordNet as an adjective. SentiWordNet assigns different positive and negative scores to these senses. The definition of sense 1 and sense 5 and their positive and negative scores are given in Table I.

TABLE I. DEFINITION WITH THEIR SENSE SCORE

Sense Id	Definition	Positive Score	Negative Score
#1	tending to become more severe or wider in scope	0.375	0.5
#5	characterized by energetic activity	0.5	0.125

Using the score of the first sense will result in drifting the overall score towards negative polarity. A lot of work has been done already on word sense disambiguation (WSD). However, WSD has been less investigated in the context of sentiment analysis. This paper investigates the effect of integrating WSD into a SentiWordNet-based approach for Sentiment Polarity classification and compares its performance with earlier reported knowledge-based approaches based on SentiWordNet that do not use sense disambiguation [1] and supervised machine learning classifiers [6-7].

Pamungkas & Putri [4] attempted to resolve the ambiguous issue of knowledge-based sentiment analysis by adopting several path-based semantic relatedness approaches. They reported improved performance on the Google Play Store dataset 1 after disambiguation. However, the path-based approach for disambiguation is computationally expensive and may not be suitable for all sentiment analysis applications. Hung and Chen [6] presented three WSD strategies for sentiment analysis tasks and observed improvement in performance on experimental investigation. Unlike the work reported by Hung and Chen [6] the present work does not attempt to revise SentiWordNet into a domain-oriented lexicon. Seifollahi and Shajari, [7] proposed a novel method of word sense disambiguation and used it to analyze the sentiment of news headlines in their system for predicting FOREX market prediction. The inclusion of WSD improves accuracy from 83.33% to 91.67%. Adverse to the conclusions of the investigation conducted by Hung and Chen[6], Seifollahi, and Shajari [7] the focus of the present work is not to propose WSD methods specific to SA but to assess how disambiguation affects the performance of sentiment analysis. Hence, we use the widely known Lesk algorithm for disambiguation and integrate it into a Lexicon-based approach to classify movie reviews into positive or negative classes based on the sentiment polarity. This classification is achieved using the overall polarity score of reviews. The polarity score is assigned using SentiWordNet. The evaluation is done on one of the benchmark datasets, namely the Cornell Movie review dataset (also known as Polarity Dataset V2.0). The dataset is publicly available for research and is widely used in earlier published works on sentiment analysis. The choice of the dataset makes it possible to compare the results with earlier reported SA work on movie reviews. The results of the proposed method henceforth referred to as Sense Enriched Lexicon-based Approach (SELSA), are compared with SentiWordNet-based SA methods that do not use WSD and supervised machine learning methods, namely Naïve Bayes (NB), Logistic Regression (LR) and Support Vector Machine (SVM). Both NB and SVM, supervised learning models have been investigated by a number of researchers on Movie Review Dataset including Pang et al. and Patel & Chhinkaniwala [8,9] using a wide variety of features. Their results reveal that SVM performs better than other models with its accuracy being highly dependent on the feature selection.

Hence, to cope with the variation in the randomly selected training and test dataset and due to additional pre-processing steps applied (stemming), we re-investigate SVM as well as NB and Logistic Regression learning models using three different features-Word2vec, Hashing Vectorizer, and Bi-gram and compare the performance of SELSA with the best performing feature-classifier combination. The important highlights of the present work are:

- Integration of computationally simple WSD method in a Lexicon-based setting for sentiment polarity classification
- Investigation of various features for supervised sentiment polarity classification
- Evaluation of the WSD-based SA approach on the benchmark Polarity dataset
- Comparison of the proposed approach with Lexicon-based as well as supervised machine learning based methods for SA

The rest of the paper is organized as follows: The second part deals with related work. In the third section, the proposed methodology is discussed. The fourth segment covered a discussion of machine learning models. The fifth section includes the dataset and the assessment metrics. In the sixth part, the present work is evaluated and compared to previous research. In the final section, the conclusions are presented.

II. RELATED WORKS

Earlier reported work in sentiment analysis attempt to classify word, sentence, or document based on positive or negative based on sentiments being expressed. Most of these works consider words as the basic sentiment-bearing unit. Word-level sentiment analysis task can be viewed as a two-step process, namely subjectivity analysis and orientation detection. Subjectivity analysis identifies subjective words, whereas orientation detection determines the semantic orientation of these words [10, 11]. The sentence or document-level sentiment classification is achieved by aggregating the sentiment of words or phrases appearing in it. The existing approaches to sentiment analysis can be broadly categorized into knowledge-based and machine learning approaches.

Knowledge-based methods, also known as Lexicon-based methods, use predefined collections of sentiment words and predict overall sentiment polarity based on the occurrences of the sentiment words in it [3, 12]. Some of the early works

[1] https://www.researchgate.net/publication/310269794_Dataset_for_Sentiment_Analysis_from_Google_Playstore_and_Twitter

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6607
***Article Received**: 19 February 2023 **Revised**: 04 April 2023 **Accepted**: 22 April 2023*

consider adjectives as strong indicators of sentiments and attempt to classify adjective words like “good”, “bad”, “happy”, etc. as positive and negative [11]. The work involving subjective nouns and verbs includes Riloff et al. and Kim & Hovy [13, 14]. A useful lexical resource that has been widely used in knowledge-based sentiment analysis research is SentiWordNet [3]. It assigns objectivity, positivity, and negativity score in the range of 0.0 to 1.0 to each WordNet synset. Knowledge-based methods are fast but assume the existence of a Lexicon. In Agrawal & Siddiqui [1] the authors propose a method for sentiment polarity analysis that applies heuristics to the results obtained using SentiWordNet to handle context-dependent sentiment expressions. They use the score of synsets of the same parts of speech instead of the score of all the synsets of a word listed in SentiWordNet. The heuristics were used to handle negation and compound sentences joined using conjunctions like “but”, “and”, “however”, etc. which gives important clues about the polarity of subjective words. An extension of this work is reported in Agrawal & Siddiqui [2] which focuses on generating ratings based on individual features. The results are reported for four different features - story, cast, music, and cinematography. They also experimented with these features by considering their synonyms from WordNet and experienced slight improvement. However, no disambiguation was done and the score of the most frequent sense was used in computing the polarity score.

The development of machine learning methods in natural language processing and information retrieval and the availability of data due to the expansion of the Internet has made machine learning approaches quite popular for sentiment analysis tasks [3, 15]. Machine learning classifiers such as K-NN, Decision Tree (DT), Naive Bayes (NB), Support Vector Machine (SVM), and Maximum Entropy (ME) have been widely used in SA tasks [8, 9, 16]. The accuracy of these algorithms greatly depends on the features selected. The pioneering work using a machine learning classifier to detect the polarity of movie reviews was done by Pang et al. [8]. They evaluated three supervised machine learning algorithms, namely NB, ME, and SVM using n-gram features. They considered only those uni-grams which occur at least four times in their corpus comprising 1400 documents and additionally added a NOT_ tag to every word appearing between a negation word. For bi-grams, the cut-off on frequency was 7 and NOT_tag was not used. All three-classifier performed well in comparison to the baseline. In terms of relative performance, SVM performed best and NB worst. The maximum accuracy of 82.9% was observed using uni-gram with an SVM classifier using binary vector representation.

Tripathi and Naganna [16] evaluated the performance of NB and SVM using n-gram (n=1, 2, 3 & 4) features on Polarity Dataset v2.0. SVM classifier with bi-gram feature outperformed all other feature-classifier combinations. This is unlike Pang et al. [8] who observed better performance with the bi-gram feature. Uni-gram feature performed comparably to the bi-gram. However, a drop in performance was observed using higher-order n-grams.

Patel and Chhinkaniwala [9] evaluated the performance of ME, NB, and SVM on two benchmark datasets, one of them being Movie Review Dataset and observed that SVM performs better than ME and NB.

In Dasilvaa et al. [17] different combinations of bag-of-words (BoW), feature hashing, and lexicons have been investigated to boost the accuracy of an ensemble classifier. The results are compared with stand-alone classifiers.

Sazzed and Jayarathna [18] proposed SSentiA (Self-supervised Sentiment Analyzer), a self-supervised hybrid methodology for sentiment classification from unlabeled data that combines an ML classifier with a lexicon-based method. They start by introducing LRSentiA (Lexical Rule-based Sentiment Analyzer), a lexicon-based method for predicting the semantic orientation of a review and its confidence score. They construct highly accurate pseudo-labels for SSentiA that use a supervised ML approach to improve the effectiveness of sentiment classification for less polarized and complex reviews using the confidence scores of LRSentiA. They compare the performance of LRSentiA and SSSentA against that of existing unsupervised, lexicon-based, and self-supervised approaches.

Gupta & Joshi [19] presented a hybrid approach to enhance the classification performance which uses a SentiWordNet-based feature vector as input to an SVM classifier. They attempted to model negation during the generation of a SentiWordNet-based feature by shifting of score in the presence of negative cue words. The generated feature vectors were used to train the SVM classifier. The experimental investigation done on the Twitter dataset from SemEval-2013 task 2 competitions demonstrates an improved performance. A comprehensive review of various techniques for sentiment analysis can be found in Pang & Lee and Medhat et al. [15, 2 0]. More recently, Birjali et al. and Wankhade et al. [21, 22] summarized various sentiment analysis methods, their challenges, and their applications.

Khan et al. [23] created a sentiment dictionary, SentiMI, and proposed a method for sentiment classification using Mutual Information (MI) score. In order to compute MI, they used SentiWordNet 3.0 as a labeled dataset and computed the MI value for each synset-POS combination. Their SentiMI-based algorithm obtained an accuracy of 84% on the Cornell Movie Review dataset.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6607
***Article Received**: 19 February 2023 **Revised**: 04 April 2023 **Accepted**: 22 April 2023*

III. METHODOLOGY

This section discusses the proposed Sense Enriched Lexicon-based approach for sentiment analysis (SELSA). The various features being investigated for supervised sentiment polarity classification are also discussed.

A. Sense Enriched Lexicon-based Approach for Sentiment Analysis (SELSA)

Unlike most of the existing Lexicon-based Sentiment analysis based on SentiWordNet, the proposed algorithm uses Lesk's algorithm to identify the correct sense of the subjective words. Instead of using the score of the dominant sense the score of the contextually appropriate sense as identified by the WSD algorithm is used in determining the sentiment polarity. The steps of the proposed algorithm are listed below:

1. Pre-process the text to remove stop words and to reduce morphological variants to stem using NLTK ((Natural Language ToolKit).
2. Assign Part of Speech tag to each word
3. Find ambiguous words using WSD package of NLTK.
4. Prepare context vector for disambiguation. A window size of ±3 words is used for creating context vector.
5. Determine the sense of the ambiguous words appearing in the review.
6. Retrieve the polarity score (positive and negative) of words from the SentiWordNet of the sense identified in step 5.
7. Compute positive and negative score of the test instance by aggregating positive and negative sentiment score of words appearing in it.
8. The class with maximum score is the winner class.

B. Classifier Selection for Comparison

As discussed earlier, supervised machine learning classifiers have been investigated by a number of researchers on a wide variety of features for sentiment analysis task [8, 9, 16]. In all these works, SVM performed better on Movie Review dataset. Pang et al. [8] observed the best performance with SVM classifier using uni-gram feature on movie review dataset. In a study conducted by Liu et al. [24]the performance of SVM classifier with bi-gram feature was found better than ME and NB and also better than SVM using uni-gram feature. This is unlike Pang et al. [8] who obtained maximum accuracy by SVM using uni-gram feature. Patel & Chhinkaniwala [9] investigated ME, NB and SVM classifier using different combinations of bag-of-words (BoW), feature hashing and lexicons and found that SVM outperformed all other classifiers. Considering the findings of these works, we choose SVM for performance comparison. However, to re-confirm these results and in an attempt to find better feature-classifier combination, we investigate the performance SVM as well as LR and NB using bi-gram and two additional features namely, word2vec and hashing vectorizer.

1) Bi-Gram: N-grams have been widely used in NLP applications. Example includes POS tagger, information retrieval, topic categorization, etc. It refers to a subsequence of n consecutive units. The unit may refer to character or word depending on application. Bi-gram is a specific case of n-gram corresponding to n=2. In this work, stop words are first removed and then pair of adjacent words are extracted.

2) Word2Vec: Word2vec is one of the popular feature extraction techniques for computing word embedding. Word2vec groups similar words and makes highly accurate guesses about meaning of words based on the context. It can be easily applied to any dataset of natural language. There are two underlying architecture models for word2vec: Continuous Bag-of-Words (CBoW) and skip gram model. Skip Gram model is used to predict context words using target word whereas Continuous Bag-of-Word model is used to predict target word from distributed representation of context. CBoW is computationally less expensive than skip grams and is chosen for this work.

3) Hashing Vectorizer: Hashing Vectorizer converts a collection of text documents to a matrix of token occurrences. Hashing Vectorizer maps each token to a column position in a matrix of predefined size. The hash function used in this work is murmurhash3.

IV. SUPERVISED MACHINE LEARNING MODELS

The process of developing models that are capable of carrying out specific tasks without the need for a human to explicitly train it to do so is referred as machine learning. Machine learning algorithms can be divided into three categories: (i) Supervised machine learning: Data collection is labeled so you can determine what inputs should be mapped to what outputs. There are two types of supervised machine learning model: Classification- assigns incoming data into either of the two classes. Regression- predicts a continuous-valued output. (ii) Unsupervised machine learning (iii) Reinforcement machine learning. The following are the supervised machine learning models that we utilized in our experiment to compare performance.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6607
***Article Received**: 19 February 2023 **Revised**: 04 April 2023 **Accepted**: 22 April 2023*

A. Naïve Bayes: Naive Bayes (NB) is a probabilistic classification model based on Bayes theorem. Regardless of the fact that the independence assumption does not hold in many of the real-world situation, NB classifiers works efficiently in numerous problems, e.g., spam detection or filtering, document classification, etc. NB classifier functions by computing posterior probability of each class using likelihood and class prior probability and returns the class which has the maximum posterior probability. It requires training data to compute likelihood and class prior probability.

$$c' = \text{argmax}_{c\in C}\ P(c|d) = \text{argmax}_{c\in C} P(d|c)P(c) \quad (i)$$

where, P(c/d) is the posterior probability of class given the document d, P(c) is the prior probability of the class c and P(d/c) is likelihood of document d. The likelihood is computed in terms of featuresfeaturesf1, f2,…, fn extracted from the document and using the conditional independence assumption, known as Naive Bayes assumption, as follows:

$$P(d|c) = P(f_1|c).P(f_2|c)\dots P(f_n|c) \quad (ii)$$

B. Support Vector Machine (SVM): Support Vector Machine is a supervised machine learning model. In their basic form, SVMs learn linear threshold function. For a binary classification task, it attempts to find the hyperplane that maximizes the margin between the two classes. Using kernel trick, they can be easily extended to learn various types of non-linear decision boundaries. One remarkable property of SVM is that they can generalize well in high dimensional feature space. This makes them suitable for text classification applications where the dimensionality is often very large.

C. Logistic Regression: Logistic regression is a supervised machine learning model. It models the probability that describes the possible outcome of a single trial using a logistic function. Logistic regression is designed for binary classification task and is effective for understanding the influence of various independent variables on the outcome variable. It anticipates that data is free from missing values and all predictors are independent of each other.

V. EXERIMENTAL EVALUATION

Dataset: The evaluation of the proposed algorithm is done on a publicly available polarity dataset v2.0 on Cornell's website2. This is a benchmark dataset as many researchers have used this dataset to evaluate the performance of their proposed method. The dataset comprises of 1000 positive and 1000 negative reviews which are organized into 10 different folders. Ten-fold cross validation is used in the experimental investigation of ML models.

Evaluation Measures: The metrics used in performance evaluation are accuracy, precision, recall and f-score. These measures are calculated using the confusion matrix (Table II).

TABLE II. CONFUSION MATRIX FOR PERFORMANCE MEASURE

	P (Predictable)	**N** (Not Predictable)
P (Actual)	**TP**	**FN**
N (Not Actual)	**FP**	**TN**

Where,
TP (True positive) =Number of correct positive predictions.
TN (True Negative) = Number of negative predictions identified
FP (False Positive) = Number of inaccurate positive predictions
FN (False Negative) = Number of inaccurate negative predictions.

Accuracy (A) is the proportion of accurately anticipated observation to total observations and can be calculated from the confusion matrix (Table I) using following expression:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (iii)$$

Precision (P)is the fraction of correctly classified positive samples. Positive predictive value (PPV) is another name for it (PPV). The highest precision is 1.0, while the lowest is 0.0.

$$P = \frac{TP}{TP + FP} \quad (iv)$$

Recall is calculated as true positive samples divided by sum of the true positive and false negative samples. It is also known as true positive rate (TPR).

$$R = \frac{TP}{TP + FN} \quad (v)$$

The F-score is often referred to as the F-Measure or F1-Score. It's a precision and recall harmonic mean.

$$F - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (vi)$$

[2]http://www.cs.cornell.edu/people/pabo/movie-review-data/review_polarity.tar.gz.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6607
***Article Received**: 19 February 2023 **Revised**: 04 April 2023 **Accepted**: 22 April 2023*

VI. EXERIMENTS AND RESULTS

Sense-enriched SentiWordNet-based method is evaluated on Movie Review Dataset. The classification accuracy resulting from this approach is shown in Table III and is compared with the work reported in Table IV. The research reported by Agrawal & Siddiqui [1] uses SentiWordNet whereas the work by Khan et al. [23] uses a sentiment dictionary, SentiMI, built using mutual information calculated for each synset-POS combination of SentiWordNet. Both these works use a dictionary-based approach for classification and are evaluated on the same dataset. This justifies the comparison. The proposed method outperforms both with an average accuracy of 96.25. This clearly demonstrates that disambiguation helps in identifying correct polarity of reviews resulting in an improved accuracy.

TABLE III. ACCURACY OF THE SENSE ENRICHED SENTIWORDNET-BASED APPROACH

Positive	Negative	Average
98.7	93.8	96.25

TABLE IV: COMPARISON OF THE PROPOSED APPROACH WITH AGARWAL & SIDDIQUI [1] AND KHAN ET AL. [23]

Method	**Positive**	**Negative**	**Average Accuracy**
Agarwal & Siddiqui [1]	90.2	80.7	85.45
Khan *et al.*[23]			84.0
SELSA	98.7	93.8	96.25

In order to compare the performance of the proposed approach with the supervised classifier, test runs are first conducted for NB, LR and SVM classifier using bi-gram, hashing vectorizer and word2vec features. The performance of these classifier is evaluated on movie review dataset. Each classifier is trained on Movie review dataset and a confusion matrix is created. The evaluation results are shown in Table V. All results reported are the average 10-fold cross-validation results. Figure 1 graphically compares their accuracy. As shown in Table V, hashingvectorizer and bigram features perform better than word2vec for all the three classifiers. The overall best accuracy of 91.9% is achieved by SVM classifier with bi-gram feature. The percentage accuracy reported in [8, 9] on the same dataset using SVM is 82.9 and 86.4 respectively. This difference is due to the difference in the feature selection process and the pre-processing steps. Patel & Chhinkaniwala [9] eliminated the stop words before extracting bi-grams.

Pang et al. [8] used only those uni-grams in experiment which has an occurrence frequency ≥ 4. In addition to stop word removal, we have also applied stemming during pre-processing. Table VI compares the performance of SELSA with the best performing feature-classifier combination investigated in this work (SVM using bi-gram feature) and the earlier reported results on the same dataset.

TABLE V: PERFORMANCE RESULTS OF NB, LR & SVM (%)

Features→		**W2V**				**HV**				**Bi-Gram**			
Performance metrics		**A**	**P**	**R**	**F**	**A**	**P**	**R**	**F**	**A**	**P**	**R**	**F**
Models	**NB**	66.3	65.9	65.9	66.7	75.0	87.0	82.0	84.0	77.6	74.5	72.0	77.0
	LR	79.9	78.6	78.0	78.0	82.3	80.5	80.5	80.0	84.5	87.0	82.0	84.0
	SVM	87.0	89.0	91.0	90.0	89.0	89.0	86.0	88.0	**91.9**	89.0	90.0	90.0

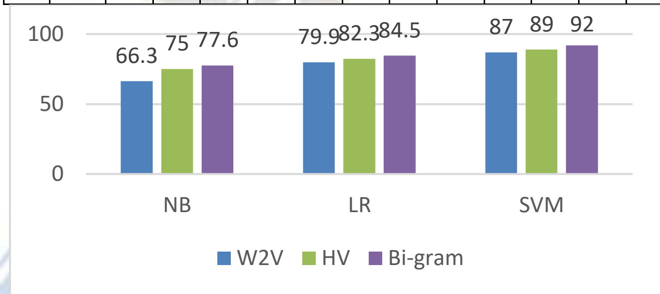

Figure 1. Performance comparison of supervised ML models using different features

TABLE VI. COMPARISON OF SELSA WITH SVM USING BI-GRAM FEATURE, PANG *ET AL.* [8] AND (PATEL & CHHINKANIWALA [9]

Method	**Positive**	**Negative**	**Average Accuracy (%)**
SELSA	98.7	93.8	96.25
SVM (using Bi-gram feature)	94.6	89.3	91.95
Pang *et al.* [8]			82.9
(Patel& Chhinkaniwala [9]			86.4

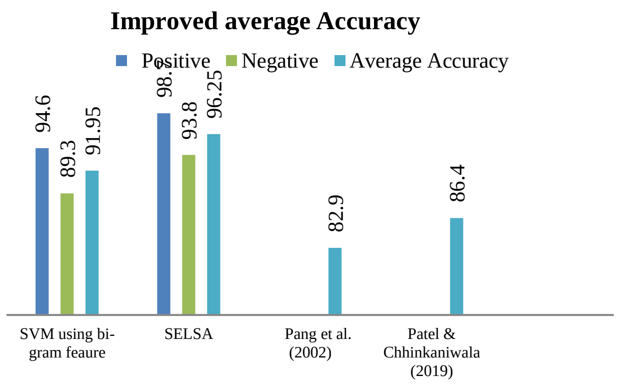

Figure 2. Graphical comparison of SELSA with best performing supervised classifiers reported in [8,9]

As shown in Table III, SVM with bi-gram features performs better on Movie Review dataset as compared to lexicon-based approach reported in Agrawal & Siddiqui [1] and Khan et al. [23]. However, after disambiguation the performance of the proposed Lexicon-based approach, SELSA, is improved. It achieves an average accuracy of 96.25% which is significantly better than the state-of-the-art supervised machine learning models available in existing literature. This is due to the fact that SentiWordNet assigns different scores to different senses of a word. After disambiguation, the score of the contextually appropriate sense is used in computing positive and negative score. This results in correct classification of some of the earlier misclassified instances resulting in an improvement in overall accuracy.

VII. Conclusion

A Lexicon-based approach for sentiment polarity classification is proposed in this work which attempts to improve the accuracy using word sense disambiguation. The proposed approach, called SELSA, is evaluated on movie review dataset which has been used earlier by many researchers. This enables comparison of proposed approach with earlier published works. The observed accuracy using SELSA is 96.25 which is significantly better than lexicon-based approaches existing in the available literature Agrawal & Siddiqui, (2009) and Khan et al. (2016). The performance of SELSA is also compared with supervised machine learning models. The best performing supervised classifier reported in existing literature is SVM. However, the accuracy is dependent on feature selection. Hence, this work investigates three supervised approaches for sentiment polarity identification of movie review sentiment using bi-gram feature which reportedly performed better in earlier work. Additionally, two more features were investigated – Word2Vec and hashing vectorizer. The performance of SVM using bi-gram feature was found better than all other feature-classifier combination. The performance of the proposed algorithm is compared with this mod el and two other supervised classifiers reported by Devitt & Ahmad, (2007), Patel & Chhinkaniwala (2019). SELSA performs significantly better than all these models in terms of accuracy. This proves the effectiveness of the proposed approach.

References

[1] S. Agrawal and T. Siddiqui, "Using syntactic and contextual information for sentiment polarity analysis", ICIS, Seoul, Korea, ACM, 978-1-60558-710-3/09/11, (2009).

[2] S. Agrawal and T.J. Siddiqui, "Feature based star rating of reviews: A knowledge-based approach for document sentiment classification", International Journal of Hybrid Information Technology 5 (4), 95-110, (2012).

[3] A. Devitt, and K. Ahmad, " Sentiment Polarity Identification in Financial News: A Cohesion-based Approach", Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, pp. 984–991, (2007).

[4] E. W. Pamungkas and D.G.P. Putri, "Word sense disambiguation for lexicon-based sentiment analysis", In Proceedings of the 9th International Conference on Machine Learning and Computing, 442-446, (2017).

[5] S. Tofighy and S. M. Fakhrahmad, "A proposed scheme for sentiment analysis: Effective feature reduction based on statistical information of SentiWordNet", Kybernetes, Vol. 47 No. 5, pp. 957-984. https://doi.org/10.1108/K-06-2017-0229, (2018).

[6] C. Hung, and S. J. Chen, "Word sense disambiguation-based sentiment lexicons for sentiment classification", Knowledge-Based Systems, Volume 110, Pages 224-232, https://doi.org/10.1016/j.knosys.2016.07.030, (2016).

[7] S. Seifollahi and M. Shajari, "Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to FOREX market prediction", Journal of Intelligent System, 52, 57–83, https://doi.org/10.1007/s10844-018-0504-9, (2019).

[8] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, ACL, pp. 79-86, (2002).

[9] N. V. Patel and H. Chhinkaniwala, "Investigating Machine Learning Techniques for User Sentiment Analysis", International Journal of Decision Support System Technology, 11(3), (2019).

[10] P.D. Turney, and M.L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association", ACM Transactions on Information Systems, 21(4), pages 315–346, (2003).

[11] V. Hatzivassiloglou and K. R. McKeown, "Predicting the Semantic Orientation of Adjectives", Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics, Madrid, ES, pp. 174–181, (1997).

[12] M. Taboada, J. M. Brooke, K. V. Tofiloski and M. Stede, "Lexicon-based methods for sentiment analysis", Computational Linguistics, 37 (2), 267–307, (2011).

[13] E. Riloff, J. Wiebe and T. Wilson, "Learning Subjective Nouns using Extraction Pattern Bootstrapping", Proceedings of CONLL-03, 7th Conference on Natural Language Learning, Edmonton, CA, pp. 25–32, (2003).

[14] S. M. Kim, and E. Hovy, "Determining the Sentiment of Opinions", Proceedings of COLING-04, 20th International Conference on Computational Linguistics, Geneva, pp. 1367–1373, (2004).

[15] B. Pang, and L. Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, vol. 2, pp. 1–135, (2008).

ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6607
***Article Received**: 19 February 2023 **Revised**: 04 April 2023 **Accepted**: 22 April 2023*

[16] G. Tripathi and S. Naganna, “Feature selection and classification approach for sentiment analysis”, Machine Learning and Applications, An International Journal, 2(2), (2015).

[17] N. F. F. Dasilvaa, E.R. Hruschkaa and E. R. Hruschka Jr, “Tweet sentiment analysis with classifier ensembles”, Decision Support Systems, Elsevier, 60, 170-179, (2014).

[18] S. Sazzed and S. Jayarathna, “A Self-supervised Sentiment Analyzer for classification from unlabeled data, Machine Learning with Applications”, Volume 4, 15 June 2021, 100026, https://doi.org/10.1016/j.mlwa.2021.100026, (2021).

[19] I. Gupta and N. Joshi, “Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic”, Journal of Intelligent System, 29(1): 1611-1625, DeGruyter, https://doi.org/10.1515/jisys-2019-0106, (2020).

[20] W. Medhat, A. Hassan and H. Korashy, “Sentiment analysis algorithms and applications: A survey”, Ain Shams Engineering Journal, Elsevier, 5(4), 1093-1113, (2014).

[21] M. Birjali, M. Kasri, and A. B. Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends”, Knowledge-Based Systems, Elsevier, Volume 226, (2021).

[22] M. Wankhade, A. C. S. Rao and C. Kulkarni, “A survey on sentiment analysis methods, applications and challenges”, Artificial Intelligence Review, https://doi.org/10.1007/s10462-022-10144-1, (2022)

[23] F. H. Khan, U. Qamar and S. Bashir, “SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection”, Applied Soft Computing, pp. 140-153, 39, (2016).

[24] S. Liu, F. Li, F. Li, X. Cheng and H. Shen, “Adaptive co-training SVM for sentiment classification on tweets”, In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, ACM, (pp. 2079-2088), (2013).