

Document-level sentiment classification using hybrid machine learning approach

Abinash Tripathy¹  · Abhishek Anand¹ · Santanu Kumar Rath¹

Received: 21 April 2016 / Revised: 16 April 2017 / Accepted: 18 April 2017 /
Published online: 10 May 2017
© Springer-Verlag London 2017

Abstract It is a practice that users or customers intend to share their comments or reviews about any product in different social networking sites. An analyst usually processes to reviews properly to obtain any meaningful information from it. Classification of sentiments associated with reviews is one of these processing steps. The reviews framed are often made in text format. While processing the text reviews, each word of the review is considered as a feature. Thus, selection of right kind of features needs to be carried out to select the best feature from the set of all features. In this paper, the machine learning algorithm, i.e., support vector machine, is used to select the best features from the training data. These features are then given input to artificial neural network method, to process further. Different performance evaluation parameters such as precision, recall, f-measure, accuracy have been considered to evaluate the performance of the proposed approach on two different datasets, i.e., IMDb dataset and polarity dataset.

Keywords Document-level sentiment analysis · Machine learning algorithm · Support vector machine (SVM) · Artificial neural network (ANN) · Performance evaluation parameter

1 Introduction

Considering the case of marketing management, the review comments on any product help not only the customer to buy the best product available in the market but also the seller, to

✉ Abinash Tripathy
abi.tripathy@gmail.com

Abhishek Anand
abhianand1093@gmail.com

Santanu Kumar Rath
skrath@nitrkl.ac.in

¹ Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela, India

know the good or bad features related to their product. These reviews are mostly available in the text format in an unstructured way and those need to be classified properly in order to provide any useful information to both customer and seller. Thus, the process of classification plays an important role in analysis of reviews [1–3]. The concept of sentiment analysis, also known as opinion mining, analyzes people’s opinion as well as emotions toward different entities such as products, organizations and their attributes [4].

In general, sentiment analysis has been investigated by various researchers mainly in three levels such as document level, sentence level and aspect level [5]. *Document level* classifies as to whether the whole opinion document expresses a positive or negative sentiment. *Sentence level* determines whether the sentence expresses any negative, positive or neutral opinion. *Aspect level* focuses on all sentiments as expressions present within given document and the aspect to which it refers. Aspect-based sentiment analysis is mainly concerned to a particular aspect of the topic or product [4]. When the analysis of the document is carried out based on aspect basis, the document may provide positive polarity, while for other aspect, it may provide the negative one. Thus, the polarity of the document depends upon the aspect considered for analysis. Again, expressions associated with sentiment are mainly the words or features which express the sentiment of the text, i.e., the adjective and adverb words [48]. Thus, during the process of analysis of the reviews the sentiment words are collected and based on the polarity values of those sentiment analysis are carried out. But as the type of the words is not fixed throughout the text, it is difficult to analyze the polarity of text. In positive–negative polarity, i.e., two class problem, the document is analyzed properly to provide the sentiment polarity of the text. During the course of this study, document-level sentiment analysis has been taken into consideration.

There are mainly three types of machine learning techniques which are very often used in sentiment analysis, i.e., supervised, unsupervised and semi-supervised learning. In the supervised learning category, the dataset is labeled and is trained to obtain a reasonable output that helps in proper decision making [6]. Unlike supervised learning, unsupervised learning process does not provide any label data; hence, they cannot be processed at ease. In order to solve the problem of processing of unlabeled data, clustering algorithms are applied [7]. The semi-supervised approach is used, where there is a small volume of labeled data for training and a large volume of unlabeled data for testing. The main purpose of semi-supervised approach is to label the large size unlabeled data with the help of labeled data whose size is small [8]. This study presents the impact of supervised learning method on labeled data.

In order to analyze sentiment reviews, it is observed that various machine learning techniques have been considered by researchers and practitioners. In this study, both support vector machine (SVM) and artificial neural network (ANN) algorithms are used in combination for classification. SVM helps to select the best features from the set of features. These features are then considered for input to ANN, which in turn helps classify the testing reviews into positive or negative group.

The contribution of the paper can be stated as follows:

- i. Two different movie review datasets IMDb [9] and polarity [2] are considered for sentiment classification. The IMDb dataset has separate data for training and testing purpose but as polarity dataset has no such separation, tenfold cross-validation technique is used for classification.
- ii. After removal of stop words and unwanted information from the training data in both datasets, the rest words are considered as features for sentiment classification. The senti-

- ment values of each words are calculated using SVM; then, for feature selection purpose, the words having sentiment values higher than $\text{mod}(0.009)$ are considered.
- iii. These selected features/words are then given input to ANN, which tests the testing data based on selected features and classifies the review dataset into either positive or negative polarity.
 - iv. During the analysis using ANN, the number of hidden nodes is kept on changing in order to find out the best possible solution. With the help of confusion matrix and other performance evaluation parameters like “precision,” “recall,” “f-measure” and “accuracy,” the performance of the proposed approach is accessed and then compared with the results obtained by different authors as reported in the literature and it is observed that the result obtained by the proposed approach performs more accurately.

The organization of the paper is as follows: The next section discusses on the articles as available on the literature and justifies the use of proposed approach. Section 3 indicates the methodologies used in sentiment classification. Section 4 informs about the setup for the present approach. Section 5 indicates the performance evaluation of proposed approach. Section 6 finally concludes the paper.

2 Literature survey

It is observed that articles of good number of authors are available in the literature that is quite rich on sentiment classification of textual reviews. The datasets which are used in this study are mostly considered to be of movie reviews i.e., Internet movie database (IMDb) [9] and polarity dataset [2]. The following subsection is concerned with the literature review on document-level sentiment classification, which is very much relevant to this study.

2.1 Document-level sentiment classification

Pang et al. [1] have undertaken the sentiment classification with three standard algorithms, i.e., Naive Bayes classification, maximum entropy classification and SVM applied over the n-gram technique. In this paper, the techniques of unigram, bigram and their combination are considered and the analysis is made on features such as frequency and presence. Pang and Lee [2] have labeled the sentences of a document as either subjective or objective. They have implemented machine learning classifier to the subjective group which prevents polarity classification from considering useless and misleading data. They have explored extraction of methods on the basis of minimum-cut formulation that provides effective way for integration of intersentence-level information with bag of words. Matsumoto et al. [11] have considered the syntactic relationship among words as a basis of document-level sentiment analysis. Frequent word subsequence and dependency subtrees are extracted from sentences, and they act as features for SVM algorithm. They have used unigram method, bigram method and combination of both as the methods for classification.

Moraes et al. [12] have made a comparison between SVM and ANN regarding document-level sentiment analysis. They have adopted feature selection technique and weighting in bag-of-words (BOW) model. Their experiment indicates that application of ANN shows superior result in comparison with SVM. Tang has suggested that learning sentiment-specific semantic representation of document is important for document-level sentiment analysis [13]. He decomposes the review documents into four constituents, i.e., word representation, sentence structure, sentence composition and document decomposition. He used three different

versions of sentiment-specific word embedding (SSWE), i.e., $SSWE_u$, $SSWE_r$ and $SSWE_h$ along with neural network for classification. They used Twitter sentiment classification on the benchmark dataset from SemEval 2013. Tu et al. [14] have used sequence and convolution kernels using different types of structures for document-level sentiment classification. They use both sequence and convolution kernels for analysis. For sequence kernels, they have used a sequence of lexical words (SW), POS tags (SP) and combination of sequence of words and POS (SWP). For dependency kernel, they have used word (DW), POS (DP) and combined word and POS settings (DWP) and similarly for simple sequence kernels (SW, SP and SWP). They used vector kernel (VK) in a bag of words as baseline. Their approach of VK + DW has shown the best result among all the proposed result. They have used polarity [2] dataset for analysis.

Liu and Chen [15] have proposed different multi-label classification on sentiment classification. According to the authors, the multi-label classification process performs the task mainly in two ways, i.e., problem transformation and algorithm adaptation. In problem transformation, the problem is transformed into multiple single-label problems. During training, the system learns from these transformed single-label data and during testing, the trained classifier makes prediction at a single label and then translates it to multiple label. Zhang et al. have proposed the classification of Chinese comments based on word2vec and SVM^{perf} [16]. Their approach mainly consists of two part. In first part, they have considered the word2vec tool to cluster similar features in order to capture the semantic features in selected domain. Then in second part, the lexicon-based and POS-based feature selection approaches are adopted to generate the training data.

Luo et al. [17] have proposed an approach to convert the text data into low-dimension emotional space (ESM). They have annotated small size words which have definite and clear meaning. They have used two different approaches for assigning weight to words by emotional tags. The total weight of all emotional tags are calculated, and based on these values, the messages are classified into different groups. Niu et al. [18] have proposed a multi-view sentiment analysis dataset including a set of image-text pair with manual annotation collected from Twitter. Their approach of sentiment analysis can be categorized into two parts, i.e., lexicon based and statistic learning. In case of lexicon-based analysis, a set of opinion words or phrases is considered which has a predefined sentiment score, while in statistic learning, various machine learning techniques are applied with dedicated textual features. Tripathy et al. [19] have used n-gram machine learning technique to perform sentiment classification using IMDb dataset. They have applied Naive Bayes, maximum entropy, SVM and stochastic gradient descent method for classification using n-gram techniques such as unigram, bigram, trigram, unigram + bigram, bigram +trigram, unigram + bigram + trigram. Among all these approaches, they find out that SVM with unigram + bigram has shown best result among other approaches.

Table 1 provides a comparative study of different approaches adopted by authors contributed to document-level sentiment classification.

2.2 Sentiment classification using hybrid machine learning approach

Section 2.1 discusses about the literature where the classification is carried out using different machine learning techniques. There are few authors who have also used hybrid methods i.e., combination of two or more machine learning algorithms to perform classification. This section discusses few literature that uses this hybridization approach.

Govindarajan [20] has proposed sentiment classification method using the hybridization of Naive Bayes and genetic algorithm. He used the concept of number of occurrence of word

Table 1 Comparison of approaches used by authors for document-level sentiment classification

Authors	Approach	Algorithm used	Obtained result (accuracy %)	Dataset used
Pang et al. [1]	Classify the dataset using different machine learning algorithms and n-gram model	Naive Bayes (NB), maximum entropy (ME), support vector machine (SVM)	Unigram: SVM (82.9), Bigram: ME (77.4), Unigram + Bigram : SVM (82.7)	Internet movie database (IMDb)
Pang and Lee [2]	Considering minimum-cut formulation method on subjective document	Naive Bayes (NB) and support vector machine (SVM)	NB: 81.5, SVMs: 65.9	Polarity dataset
Matsumoto et al. [11]	Syntactic relationship among words used as a basis of document-level sentiment analysis	SVM with N-gram techniques	Unigram: 83.7, Bigram: 80.4, Unigram+Bigram : 84.6	Internet movie database (IMDb), polarity dataset
Moraes et al. [12]	Select the terms using Information Gain ranking and then classify the reviews using NB, SVM and ANN	NB, SVM and ANN	NB:80.3, SVM : 84.1, ANN:86.5	Polarity dataset and Amazon reviews on GPS, Books and Camera
Tang [13]	Used sentiment-specific word embedding (SSWE) technique along with three different versions of Neural Network to classify reviews	Three different versions of Neural network	SSWE _u :83.37, SSWE _r : 82.6 SSWE _u : 77.33	Twitter sentiment dataset in SemEval 2013
Tu et al. [14]	Used sequence and convolution kernels using different types of structures for document-level sentiment classification	Vector Kernel (VK) with sequence and dependency kernel	VK + DW: 88.50	Polarity dataset
Liu and Chen [15]	Used multi-label classification using eleven state-of-the-art multi-label, two microblog datasets and eight different evaluation matrices on three different sentiment dictionaries	Eight different evaluation matrices	Average highest Precision: 75.5	Dalian University of Technology Sentiment Dictionary (DUTSD), National Taiwan University Sentiment Dictionary (NTUSD), Howset Dictionary (HD)

Table 1 continued

Authors	Approach	Algorithm used	Obtained result (accuracy %)	Dataset used
Zhang et al. [16]	Use word2vec to capture similar features then classify reviews using SVM ^{perf}	SVM ^{perf}	Lexicon based: 89.95, POS based: 90.30	Chinese comments on clothing products
Luo et al. [17]	Ekman Paul's research approach is used to convert the text into low dimensional emotional space (ESM), then classify them using machine learning techniques	Support vector machine (SVM), Naive Bayes (NB), decision tree (DT)	SVM: 78.31, NB: 63.28, DT: 79.21	Stock message text data (The Lion forum)
Niu et al. [18]	Lexicon-based analysis on set of opinion words or phrases. Then various machine learning techniques are used for document-level classification	Multi-model deep Boltzmann Machine	TF: 71.9	Multi-view sentiment analysis dataset
Tripathy et al. [19]	Used N-gram technique along with machine learning techniques to classify the reviews	Support vector machine (SVM), Naive Bayes (NB), maximum entropy (ME), stochastic gradient descent (SGD)	SVM: 88.94, NB: 86.23, ME: 88.48, SGD: 83.36	Internet movie database (IMDb)

i.e., the count as feature selection. The words that occur more than three times in text are only considered for further analysis. They also used best first search (BFS) approach to evaluate the classifier result. Abbasi et al. [21] have proposed an algorithm known as entropy-weighted genetic algorithm (EWGA) which is a hybridized genetic algorithm. They have combined EWGA with SVM for classification of web forums in different languages i.e., English and Arabic.

Balage Filho et al. [22] have adopted a hybrid classification using three classification approaches: rule based, lexicon based and machine learning. They also suggest that pipeline approach is used by them to extract best feature from each classifier. Their approach can be divided into two subtasks, i.e., expression-level classification and message-level classification. Jagtap and Dhotre [23] have proposed hidden Markov model (HMM) and SVM-based hybrid classification, along with advent feature extraction model. They have used voting mechanism to classify the reviews. In voting mechanism, the classifier votes for the polarity of the review. The reviews which get highest vote belong to the polarity i.e., either positive or negative. Wang et al. [24] have proposed a hybrid method based on category distinguishing ability of words and information gain is adopted for feature selection. They have performed the classification using feature selection and SVM, and their best result obtained with feature dimension equal to 3000.

Table 2 provides a comparative study of different approaches adopted by various authors using hybrid machine learning technique.

2.3 SVM as feature selection

It is observed from Sects. 2.1 and 2.2 that for sentiment classification of reviews different machine learning techniques or hybrid approaches have been considered by a good number of authors. As SVM is used in this paper for feature selection, the present section discusses about the use of SVM as feature selection.

Neumann et al. [26] have proposed four novel continuous feature selection approaches which intends to minimize classifier load. They include linear and nonlinear SVM classifier. Their focus is mainly on embedded approach, which minimizes training errors of a linear classifier. Then, they have fixed up an objective to minimize the feature selection for nonlinear SVM classifier. Lozano et al. [27] have used three different techniques for texture classification using feature selection. The subgroup-based multiple kernel learning technique performs feature selection by down-weighting or removing subsets of features having similar characteristics. Two different conventional feature selection techniques such as recursive feature elimination with different classifier and a genetic algorithm-based approach with SVM are used as decision function. These classifiers work according to tenfold cross-validation technique. Maldonado et al. [28] have used feature selection by penalizing each features that are used in dual formulation. Their approach known as Kernel penalized SVM (KP-SVM) eliminates features having low relevance for the classifier. KP-SVM also employs stopping condition which avoids the negative affect of feature elimination on classifier's performance.

Zheng et al. [29] have used feature selection for Chinese online reviews. They have used N-char-grams as well as N-POS-grams approach for potential sentimental features. Improved document frequency method is used for feature subset selection and Boolean weighting method for calculation of feature weight. Finally, Chi-square test is carried out to test significance of experimental result. They have used LIBSVM, a SVM classifier developed by Taiwan University to conduct experiment. Sharma and Dey [30] have used five different feature selection methods such as information gain (IG), gain ratio (GR), Chi statistics (CHI), Relief-F and

Table 2 Comparison of approaches used by authors for sentiment classification using hybrid machine learning technique

Author	Approach	Algorithm used	Obtained result (accuracy %)	Dataset used
Govindarajan [20]	Used hybrid of NB and GA to perform classification	Hybrid of NB and GA	Hybrid NB-GA: 93.8	Movie reviews downloaded from Bo Pang's web pages
Abbasi et al. [21]	Used entropy-weighted genetic algorithm along with SVM to classify the reviews	Hybrid of EWGA and SVM	Movie review: 91.7, US forum: 90.6 and Middle Eastern forum: 90.52	IMDb movie review dataset and US supremacist and Middle Eastern extremist group web forums
Filho et al. [22]	Used three different classification approaches, i.e., rule based, lexicon based and machine learning based to classify the reviews	Linear kernel SVM with rule-based and lexicon-based approach	Twitter2013 Dataset: 56.31, Twitter2014 dataset : 65.39	Twitter2013, SMS2013, Twitter2014, LiveJournal2014, Twitter2014Sarcasm
Jagtap and Dhotre [23]	The SVM and HMM used to classify the reviews independently and then using voting mechanism and max-min rule specify polarity to the reviews	SVM and hidden Markov model (HMM)	SVM + HMM: 82.5	Reviews collected from student reading teachers performance
Wang et al. [24]	The hybrid feature selecting method based on category distinguishing ability of words and information gain is introduced for sentiment classification	Hybrid of distinguishing ability of words and information gain	Information gain: 84.40	Corpus of 11 different car reviews is collected from the web

document frequency (DF) along with seven machine learning techniques, i.e., Naive Bayes, SVM, maximum entropy, decision tree, K-nearest neighbor, Winnow and Adaboost for sentiment analysis of movie reviews. Their result shows that GR provides a better result for feature selection and SVM performs better than all other machine learning techniques. Hardin et al. [31] have used SVM-based feature selection. They have assigned zero weight to irrelevant variables by linear SVM. After assigning zero weights to the variables, they do not play any active role in classification; thus, feature selection is carried out and finally SVM is used for classification.

Table 3 provides a comparative study of different approaches adopted by authors using SVM as feature selection technique.

2.4 Motivation for proposed approach

The above-mentioned literature survey helps to identify some possible research areas which can be extended further. The following aspects have been considered for carrying out further research.

- i. A number of authors have used the concept of part-of-speech (POS) tags or count of the occurrence of the word as a criteria for feature selection. But it is observed that the POS tag for a word is not fixed and it changes as per the context of their use. For example, the word “book” can have the POS “noun” when used as reading material, whereas in case of “ticket booking” the POS is verb. Again the occurrence of a particular word mainly depends upon the author’s writing style. Thus, it may not be suitable for using feature selection. Hence, in this paper, the sentiment values of each word are calculated and feature selection is carried out on the basis of these sentiment values.
- ii. Different authors have used a single machine learning technique for both feature selection and classification purpose. Thus, it is found out that the shortcoming of the techniques biases the final classification result. To solve this issue, in this paper SVM is used as feature selection, whereas ANN is used for sentiment classification. Hence, the bias of any algorithm does not affect the result of classification.
- iii. Most of the machine learning algorithms work on the data represented as matrix of numbers. But the sentiment data are always in text format. So, it needs to be converted to number matrix. Different authors have considered TF or TF-IDF to convert the text into matrix on numbers. But in this paper, in order to convert the text data into matrix of numbers, the combination of TF-IDF and CountVectorizer has been applied. The rows of the matrix of numbers represent a particular text file, whereas its column represents each word/feature present in that respective file which is shown in Table 4.

3 Methodology

Sentiment classification techniques may be categorized into two types, i.e., binary sentiment classification and multi-class sentiment classification [32]. In binary classification, each document d_i in D where $D = \{d_1, \dots, d_n\}$ is classified as a label C where C is a predefined category set $C = \{\text{Positive, Negative}\}$. In multi-class sentiment analysis, each document d_i is classified as a level in C^* where $C^* = \{\text{strong positive, positive, neural, negative, strong negative}\}$. The result of binary classification helps to present an indication of quality of one

Table 3 Comparison of approaches used by authors for sentiment classification using SVM as feature selection technique

Author	Approach	Algorithm used	Obtained result (accuracy %)	Dataset used
Neumann et al. [26]	Used four different feature selection techniques continuously with linear and nonlinear SVM to perform classification	Linear and nonlinear SVM	SVM: 90	Several dataset from UCI repository [52] and Colon Cancer dataset from Weston et al. [53]
Lozano et al. [27]	Used subgroup-based multiple kernel learning (MKL), for feature selection by down-weighting or eliminating subsets of features which shares similar characteristic, two different conventional feature selection techniques and a genetic algorithm-based approach with an SVM as decision function for classification	Multi-kernel learning (MKL), recursive feature elimination (RFE)	SVM-RFE: 95.8, RFE-TREE: 93.28, RFE-RF: 92.13, MKL: 89.26	Biomedical image texture dataset
Maldonado et al. [28]	Used embedded method of feature selection along with SVM RBF kernel to perform classification	SVM	DIA: 76.54, WBC: 97.5, CMA: 96.5 and LMA: 99.7	Diabetes dataset (DIA), Wisconsin Breast Cancer (WBC), Colorectal Microarray dataset (CMA) and Lymphoma Microarray dataset (LMA)
Zheng et al. [29]	Used n-char-gram and N-POS-gram technique for feature selection then use SVM to classify reviews	SVM	4-POS-gram with 150 features: 96.52, 1-char-gram with 150 features: 95	Chinese mobile phone reviews from Jing Dong, a famous electronics e-commerce site

Table 3 continued

Author	Approach	Algorithm used	Obtained result (accuracy %)	Dataset used
Sharma and Dev [30]	Used five commonly used feature selection methods in data mining research (DF, IG, GR, CHI and Relief-F) and seven machine learning-based classification techniques for sentiment analysis on online movie reviews dataset	Nave Bayes, support vector machine, maximum entropy, decision tree, K-nearest neighbor, Winnow, Adaboost	NB-GR: 90.9, SVM-GR:90.15, ME-GR: 88.85, DT-GR:75.35,KNN-GR:75.15, Winnow-GR:73.5, Adaboost-Relief-F:66.90	Polarity dataset
Hardin et al. [31]	Assign zero weight to irrelevant variables. Then, use linear SVM to classify the reviews	SVM-based feature selection	SVM: 86.5	Polarity dataset

Table 4 Matrix generated under CountVectorizer scheme

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
Sentence 1	1	1	1	0	0
Sentence 2	1	1	0	1	0
Sentence 3	1	1	0	0	1

product over others. In this paper, study on binary sentiment classification has been carried out.

3.1 Text data to numerical data conversion technique

The task of conversion of text data into numerical data is carried out using following functions such as:

1. *CountVectorizer (CV)* It converts the text document collection into a matrix of token counts [10]. This function generates a sparse matrix of the counts.
 - Calculation of CountVectorizer Matrix: Suppose we have three different documents containing following sentences.
 “Book is good.”
 “Book is average.”
 “Book is nice.”
 Matrix is generated of size 3*5, because we have 3 documents and 5 distinct features such as book, is, good, average and nice. The elements of matrix can be represented as 4.
 where each “1” in a row corresponds to the presence of a feature and 0 represents absence of a feature from particular document.
2. *Term frequency-inverse document frequency (TF-IDF)* Reflects the importance of a word in the corpus or the collection [10]. TF-IDF value increases with increase in frequency of a particular word which appears in document. In order to control the generality of more common words, the term frequency is offset by the frequency of words in corpus. Term frequency is the no. of times a particular term appears in the text. Inverse document frequency measures the occurrence of the term in all document.
 - Calculation of TF-IDF value: Suppose a movie review contains 100 words, wherein the word *Awesome* appears 10 times. The term frequency (i.e., TF) for *Awesome* is calculated as $(10/100) = 0.1$. Again, suppose there are 1 million reviews in the corpus and the word *Awesome* appears 1000 times in whole corpus. Then, the inverse document frequency (i.e., IDF) is calculated as $\log(1,000,000/1,000) = 3$. Thus, the TF-IDF value is calculated as $0.1 * 3 = 0.3$.
3. *word2vec* Word2vec was created by a team of researchers led by Tomas Mikolov at Google. Word2vec is a group of related models that are used to produce word embeddings [49]. Word2vec takes as its input a large corpus of text and produces a high-dimensional space, with each unique word in the corpus being assigned a corresponding vector in the space. Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located in close proximity to one another in the space [51].

In this paper, the CV and TF-IDF functions are used to transform the text document into a numerical vector, which is then considered as input to supervised machine learning techniques. However, it is observed that few authors have applied word2vec to transformation words in the text into vector form but in document-level sentiment analysis, the machine learning techniques need the vector representation of whole document. Thus, CV and TF-IDF are more suitable over word2vec for transformation of the text into numerical vectors.

3.2 Used dataset

In this paper, two different datasets are considered for sentiment classification. The datasets are as follows:

1. *aclIMDb dataset* The acl Internet movie database (IMDb) consists of 12,500 positive labeled test reviews and 12,500 positive labeled train reviews. Similarly there are 12,500 negative labeled test reviews and 12,500 positive labeled train reviews [9]. Apart from labeled supervised data, an unsupervised dataset is also present with 50,000 reviews.
2. *polarity dataset* The polarity dataset consists of 1000 positive reviews and 1000 negative labeled reviews [2]. Though the database contains both negative and positive reviews, it is not separately partitioned for training and testing. In order to perform the classification, the cross-validation method is being used for this dataset.

3.3 Sentiment classification technique

As mentioned in Sect. 3.2, two different datasets are used for sentiment classification. Thus, the techniques used for classification are also different in both cases.

- The IMDb dataset has separate data for training and testing [9]. Thus, the training data are used by the machine learning classifier for training purpose and on the basis of this training result, the testing data are classified. Different performance evaluation parameters along with confusion matrix are used to evaluate the performance of the classifier.
- The polarity dataset does not have separation between the testing and training data [2]. Thus, in order to perform classification, cross-validation technique is used. The cross-validation technique is used to partition the dataset into two parts, i.e., learning and validate [33]. These two sets are designed to crossover in successive rounds so that each data point must be validated. k -Fold cross-validation is the basic cross-validation method. In k -fold validation, dataset is partitioned into k different folds. Among these k -folds, $(k - 1)$ -folds are used for training and onefold is used for testing. Tenfold cross-validation is most commonly used in machine learning and classification problems, by different authors.

3.4 Application of machine learning techniques

In this paper, the following two machine learning techniques are used.

1. *Support vector machine method (SVM)* This method analyzes data and defines decision boundaries by having hyperplanes. In two category cases, the hyperplane separates the document vector in one class from another class, where the gap of separation is kept as large as possible.
For a training set with labeled pair (x_i, y_i) , $i = 1, 2, \dots$ where $x_i \in R^n$ and $y \in \{1, 0\}^l$, the SVM required to solve the following optimization problem [34] may be represented as:

$$\begin{aligned} \min_{w,b,\xi} & \frac{1}{2} W^T W + C \sum_l^{i=1} \xi_i \\ \text{subject to} & \quad y_i(w^T \phi(X_i) + b) \geq 1 - \xi_i, \\ & \quad \xi_i \geq 0. \end{aligned} \tag{1}$$

here training vector X_i is mapped to higher-dimensional space by ϕ . SVM requires input in the form of a vector of real numbers. Thus, the reviews of text file for classification may be converted to numeric value before it can be made applicable for SVM. After the text file is converted to numeric vector, it goes through a scaling process which manages the vectors in order to have the values in the range of [1, 0].

In SVM the kernel is used for pattern analysis. Mostly four different types of kernels are used for analysis in SVM. These are as follows:

- (a) Linear kernel: The linear kernel function can be represented as follows

$$K(x_i, x_j) = x_i^T x_j \tag{2}$$

where x_i and x_j are the input space vector and x_i^T is the transpose of x_i .

- (b) Polynomial kernel: For degree “ d ,” the polynomial kernel can be defined as

$$K(x_i, x_j) = \{x_i^T x_j + c\}^d \tag{3}$$

where x_i and x_j are the input space vector i.e., the features computed from training sample, “ c ” is a parameter used for the trade-off between the highest order and lowest order polynomial. Polynomial kernel with degree = 2 is mainly used in binary sentiment classification.

- (c) Gaussian radial basis function (RBF) kernel: The RBF is a real-valued function, whose value depends upon the distance from the origin. The RBF kernel can be defined as follows

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|) \quad \text{for } \gamma > 0 \tag{4}$$

where x_i and x_j are the input space vector and the value of γ can be used as $\frac{1}{2\sigma^2}$, where σ^2 is the variance of input data.

- (d) Sigmoid kernel: The sigmoid kernel can be defined as follows

$$K(x_i, x_j) = \tanh(ax_i^T x_j + b) \tag{5}$$

where x_i and x_j are the input space vector, $a > 0$ is the scaling parameter for the input data, and b is a shifting parameter that controls the threshold of mapping.

The text classification problems are mostly linearly separable [45]. The linear kernel shows better result, when there is a presence of good number of features as mapping the data to a higher space does not affect the performance. In text classification, the number of documents and features is mostly large, i.e., for IMDb dataset the total number of reviews including both training and testing is 50,000 and the features present are 159,438, while for polarity dataset the total number of reviews including both training and testing is 2000 and the number of features is 25,579. Thus, in this paper, the linear kernel is preferred to other kernels while using SVM.

2. *Artificial neural network (ANN)* Neural network method used for classification can be represented as a mapping function such as

$$F : A^d \longrightarrow A^m \tag{6}$$

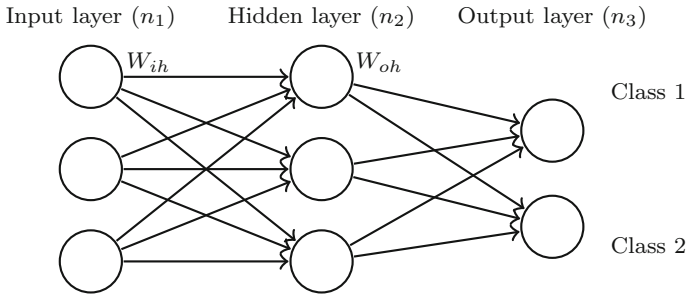


Fig. 1 A typical neural network

where “ d ”-dimensional input is submitted to network and “ m ” vector output is obtained with classification result. Following Fig. 1 shows the structure of a neural network. The input layer of neural network consists of “ d ” neurons that represent “ d ” pieces of input signal (Independent variable). The number of neurons in the hidden layer is chosen by the user. Finally the output layer consists of “ m ” number of neurons (considered as dependent variables) [35].

In the input layer, the state of each neuron is determined by input variable. For other neurons the state of neurons is evaluated using values associated with previous neurons as:

$$a_j = \sum_{i=1}^I X_i W_{ji} \tag{7}$$

where a_j is the net input of neuron j , and X_i is the output value of neuron i in previous layer. W_{ji} is the weight factor of the connection between neuron i and neuron j .

The neuron’s activity is usually determined via a sigmoid function.

$$g(a_j) = \frac{1}{1 + exp^{-a_j}} \tag{8}$$

In backpropagation technique, each iteration tries to minimize the error. The adjustment of weight is started from output layer to input layer [36]. Error correction is carried out using following function:

$$\Delta W_{ji} = \eta \delta_i F(a_i) \tag{9}$$

where ΔW_{ji} is the adjustment of weight between neuron j and i , η is the learning rate, δ_i depends on the layer, and $F(a_i)$ is the output of network “ i .” The training process is carried out till the error is minimized. After the completion of training process, the performance of ANN is tested using the input data and the result is compared using the confusion matrix.

Deep learning approaches are mainly used for analysis of the problems where there is presence of multiple types of features, i.e., for analysis of video data, images and sound all thing to be taken care [46]. Convolutional neural network (CNN) is a kind of ANN where the weight shares the network structure [47]. In CNN, continuous bag of words (CBOW) are considered for analysis. But, in this case each word is considered as a single separate unit. So, when the continuous words are considered they change the accuracy. Again, ANN is a special type of deep learning method which is mainly used when all the features are of the same time. In this paper, while classifying the reviews, the features obtained are all of the same type and as each word is considered as a separate feature and combining them impacts

Table 5 Confusion matrix

	Correct Labels	
	Positive	Negative
Positive	TP (true positive)	FP (false positive)
Negative	FN (false negative)	TN (true negative)

the accuracy result. Thus, ANN is preferred to CNN and most of the general deep learning techniques.

3.5 Performance evaluation parameters

Confusion matrix also known as contingency table is typically used in supervised machine learning techniques in order to allow visualization of performance of algorithm. From classification point of view, its element such as true positive (TP), false positive (FP), true negative (TN), false negative (FN) are used to compare label of classes [37]. True positive represents the reviews that are positive also classified as positive by the classifier, whereas false positive are positive reviews but classifier classifies it as negative. Similarly, true negative represents the reviews which are negative also classified as negative by the classifier, whereas false negative are negative reviews but classifier classifies it as positive (Table 5).

The elements of confusion matrix can be used to find the values of some of the evaluation parameters such as precision, recall and accuracy that indicates the performance of classifier.

1. Precision: It measures the exactness of the classifier result. It is the ratio of number of examples correctly labeled as positive to total no. of positively classified example.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

2. Recall: It measures the completeness of the classifier result. It is the ratio of total no. of positively labeled example to total examples that are truly positive.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

3. Accuracy: It is the most common measure of classification accuracy. It can be calculated as the ratio of correctly classified example to total number of examples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

4 Proposed approach

The IMDb and polarity datasets are preprocessed in order to remove the stop words and unwanted information from dataset. The processed textual data are then used to transform into matrix of numerical vectors using vectorization techniques. Further, the dataset goes through a feature selection step, which selects the features depending upon some conditions. The dataset is classified based on the feature selected. Stepwise elaboration of the approach is described in Algorithm 1.

Algorithm 1 Algorithm for the proposed approach

Input: Two movie review datasets IMDb and polarity are considered for classification.

Output: Reviews classified into two polarity groups i.e., positive and negative.

- 1: Dataset analysis is carried out to determine whether the general approach of training and based on that testing is carried out or K-fold cross-validation approach is used.
- 2: Preprocessing of the reviews are carried out to remove unwanted information such as stop words, numerical and special characters, URL and HTML tags, and finally stemming is carried out to obtain the root word using Porter Stemmer.
- 3: The preprocessed text reviews are then transformed into numerical vectors using CountVectorizer and TF-IDF approach.
- 4: **for** Each Feature f_i in $review_i$ the sentiment score is calculated using SVM **do**
- 5: **if** The sentiment score of the feature is greater than mod (0.009) **then**
- 6: Selected the feature for next level of analysis
- 7: **else**
- 8: The feature is rejected
- 9: **end if**
- 10: **end for**
- 11: The selected features are then given input to ANN for classification. Using varying number of hidden nodes, the accuracy of the proposed system is found out until the best possible result is obtained.

The detailed description of Algorithm 1 is presented as follows:

Step 1. The two datasets used for classification are as follows:

- The IMDb dataset consists of 12500 positive and 12500 negative reviews for training, and the same amount of reviews is there for testing [9].
- The polarity dataset consists of 1000 positive and 1000 negative reviews for analysis, where the reviews are not separated into test and training reviews [2].

Step 2. It is observed that the text reviews consist of absurd information, which need to be removed from the original reviews before it is considered for classification. The absurd information is

- Stop words: Stop words do not play any role in determining the sentiment and thus may be removed. The list of the stop words are collected from the site “<http://norm.al/2009/04/14/list-of-english-stop-words/>,” and then a list of those words being created and if the words appear in the text reviews, then they are removed considered them as the stop words.
- Numeric and special character: In the text reviews, there are different numeric (1,2,...,5, etc.) and special characters (@, #, \$,%, etc.) present, which do not have any effect on the analysis, but they create confusion while conversion of text file to numeric vector.
- URL and HTML tags: This information also needs to be removed as they do not play any role in finding out the sentiment.

After the absurd information is removed, the stemming process is carried out, i.e., the process of getting the root word from any word. For example, the root word for reading is read. For the stemming purpose, PorterStemmer tool is used. It is used to remove the common morphological and inflexional endings from words in English [50].

Step 3. After the preprocessing of text reviews, they need to be arranged into a matrix form of numeric vectors. The algorithms for conversion of text file to numeric vectors are as follows:

- CV: It converts the text reviews into a matrix of token counts. It implements both tokenization and occurrence counting.
 - TF-IDF: It suggests the importance of the word to the document and to the whole corpus. Term frequency informs about the frequency of a word in a document, and IDF informs about the frequency of the particular word in whole corpus.
- Step 4. After the text reviews are converted to numeric vectors, this information is then considered for the process of feature selection using SVM algorithm. The steps for feature selection for different datasets are shown as follows:
- (a) The IMDb dataset has 12,500 positive reviews and 12,500 negative reviews for training and the same amount of reviews present for testing purpose. For the purpose of feature selection, only the training data are considered. The polarity dataset has 1000 positive and 1000 negative reviews. As tenfold cross-validation technique is used for classification, 900 positive and negative reviews are considered as input for feature selection. As the step of feature selection is mainly carried out on training data.
 - (b) After the preprocessing stage, when the unwanted words or information is removed, rest of the words are then considered as feature. After preprocessing, the total number of features obtained from **IMDb dataset** is **159438** and from **polarity dataset** is **25579 features**.
 - (c) Then a matrix is generated, where the row specifies the file and the column specifies the feature with its occurrence.

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix}$$

Each element $x(i, j)$ represents the occurrence of feature j in review i , and α_i is a random variable multiplied with the feature.

$$\begin{bmatrix} \alpha_1 x_{1,1} + \alpha_2 x_{1,2} + \alpha_3 x_{1,3} + \dots + \alpha_{25579} x_{1,25579} \\ \alpha_1 x_{2,1} + \alpha_2 x_{2,2} + \alpha_3 x_{2,3} + \dots + \alpha_{25579} x_{2,25579} \\ \alpha_1 x_{3,1} + \alpha_2 x_{3,2} + \alpha_3 x_{3,3} + \dots + \alpha_{25579} x_{3,25579} \\ \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ \alpha_1 x_{2000,1} + \alpha_2 x_{2000,2} \dots + \alpha_{25579} x_{2000,25579} \end{bmatrix}$$

As the review is supervised, it is known that whether the sum of the product of α_1 and $x_{i,j}$ is turning out to be positive or negative. If the sum total of the product is positive, then its review is considered to be of positive polarity or else it is of negative polarity. As the value of α_1 is considered at random, sometimes the polarities do not match; in that case another set of α_1 is considered. These α_1 values finally show the polarity value of the respective features.

Table 6 Result obtained using different number of hidden nodes on IMDB dataset

Number of hidden neurons	Precision		Recall		Accuracy
	Positive	Negative	Positive	Negative	
100	.67	.76	.73	.70	.714
200	.82	.83	.83	.83	.83
300	.84	.83	.83	.84	.84
400	.88	.86	.86	.88	.87
500	.88	.89	.89	.88	.884
600	.99	.91	.92	.99	.95
700	.92	.9	.9	.92	.91
800	.88	.88	.86	.88	.87
900	.79	.91	.9	.81	.851
1000	.89	.79	.81	.88	.84
2000	.88	.78	.89	.87	.83
3000	.82	.83	.83	.83	.83
4000	.67	.76	.73	.7	.713
5000	.81	.6	.67	.76	.705

The bold value shows the highest accuracy obtained by the proposed method on IMDB and Polarity dataset

- (d) Thus, for each of features, the polarity value is obtained, but all words do not affect the polarity of review in the same order. In this present paper, the features are selected whose sentiment value is greater than $\text{mod}(0.009)$. Thus, the set of features is reduced to 19,729 from 159,438 for IMDB dataset and to 3199 from 25,579 for polarity dataset.

Step 5. After the feature selection is complete, then the input data are considered for testing. For classification, the tenfold cross-validation is being used i.e., 90% of the review are used for training i.e., already being done. Then for the testing of the rest, 10 % of reviews are carried out. The result is being analyzed using various performance evaluation parameters like precision, recall and accuracy. The number of hidden neurons depends upon the user to obtain improved value of accuracy; so, in this paper, the numbers of hidden nodes are considered to be order of 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000 and 5000. The input matrix considered for IMDB dataset is of size $25,000 \times 19,729$, and the output has two neurons, i.e., either positive or negative. The input matrix considered for polarity dataset is of size 1800×3199 , and the output has two neurons, i.e., either positive or negative.

Following Tables 6 and 7 show the result obtained using ANN classification on IMDB dataset and polarity dataset, respectively.

From Table 6 and Fig. 2, it can be observed that the hidden nodes of ANN are kept on changing until to best possible result is obtained for IMDB dataset. It is observed that the accuracy of the proposed system is in a increasing mode till the 600 hidden node. After 600 nodes, the accuracy of the system is stable or in a decreasing mode due to the overfitting of the machine learning algorithm.

From Table 7 and Fig. 3, it can be found that the accuracy of the proposed system is in an increasing mode up to a specific number of hidden nodes i.e., up to 500 hidden nodes the accuracy of the system increases and after that the accuracy of the system is either stable or decreases. Thus, the hidden nodes are kept up to 5000. The accuracy of the system decreases due to overfitting of the machine learning algorithm.

Table 7 Result obtained using different number of hidden nodes on polarity dataset

Number of hidden neurons	Precision		Recall		Accuracy
	Positive	Negative	Positive	Negative	
100	.689	.714	.732	.67	.701
200	.869	.865	.863	.87	.867
300	.867	.857	.855	.869	.862
400	.902	.914	.915	.901	.908
500	.97	.958	.957	.97	.964
600	.958	.967	.967	.958	.963
700	.934	.931	.931	.934	.933
800	.931	.947	.948	.93	.939
900	.928	.933	.933	.928	.93
1000	.902	.899	.899	.902	.9
2000	.905	.906	.906	.905	.905
3000	.909	.914	.914	.909	.911
4000	.882	.866	.863	.885	.874
5000	.83	.84	.842	.828	.835

The bold value shows the highest accuracy obtained by the proposed method on IMDb and Polarity dataset

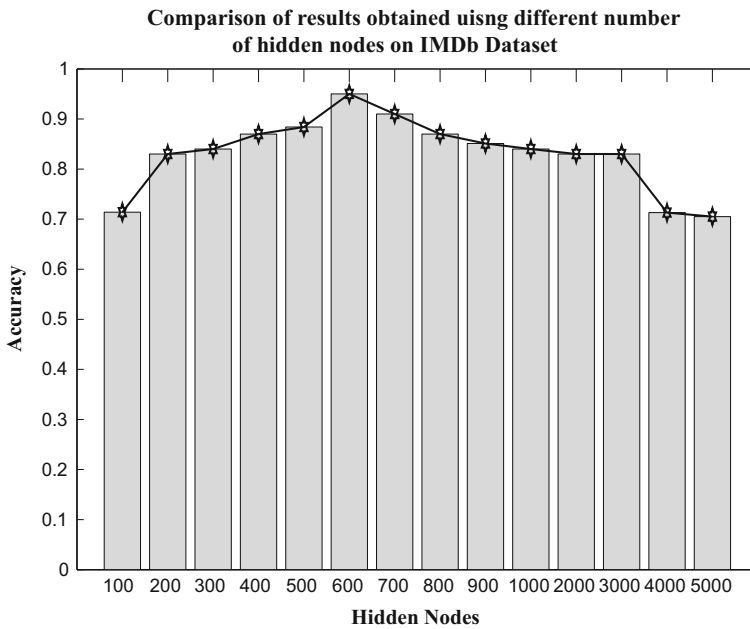


Fig. 2 Result obtained using different number of hidden nodes on IMDb dataset

5 Performance evaluation

Table 8 and Fig. 4 show the comparison of accuracy values using the proposed approach with other approaches as available in literature using IMDb dataset. It can be observed from both figure and table that most of authors have preferred either NB, SVM or combination

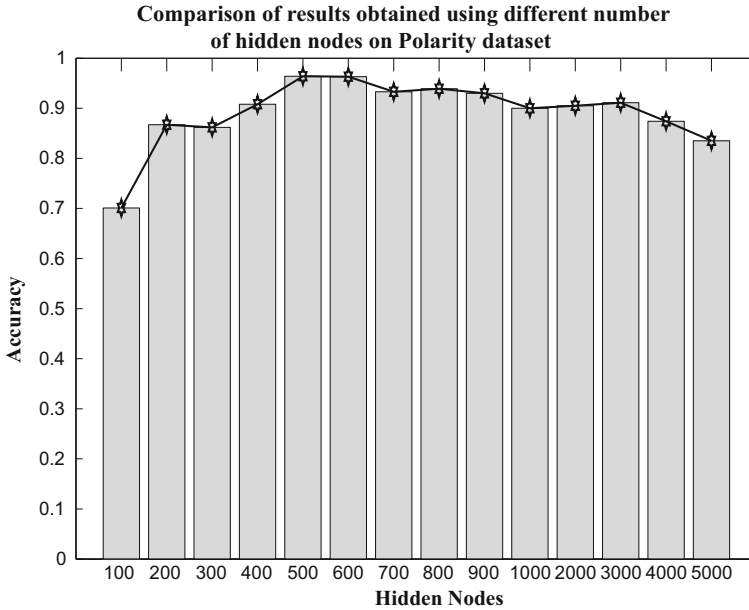


Fig. 3 Result obtained using different number of hidden nodes on polarity dataset

Table 8 Comparative result obtained by different literature using IMDb dataset

Author	Machine learning techniques used	Accuracy
Pang et al. [1]	Naive Bayes (NB), SVM	NB = 0.815, SVM = 0.659
Salvetti et al. [38]	NB	NB = 0.796
Mullen and Collier [39]	SVM	SVM = 0.86
Beineke et al. [40]	NB	NB = 0.659
Matsumoto et al. [11]	SVM	SVM = 0.883
Proposed approach	Hybrid of SVM and ANN	SVM + ANN = 0.95

of them for classification. But, in the proposed approach of combining SVM and ANN, the result obtained by the proposed approach is found out to be better than that results obtained by other authors. Different authors have considered all words for classification, but when feature selection is carried out, the obtained result is found out to be more accurate.

Table 9 and Fig. 5 show the comparison of accuracy values using the proposed approach with other approaches as available in the literature using polarity dataset. It is observed from both table and figure that most of the authors have preferred NB, SVM for classification. Mores et al. [12] have used ANN along with NB and SVM for classification. It is evident from the figure and table that the proposed approach has shown value with better accuracy as compared to result of other literature. As there is no separation between testing and training data in polarity dataset, cross-validation technique is used by different authors for classification. The authors have used tenfold cross-validation technique where 90% of dataset are used for training and rest 10% are used for testing. Again, as in the proposed approach, instead of using all feature which is done by authors, the features with best sentiment values are considered for classification.

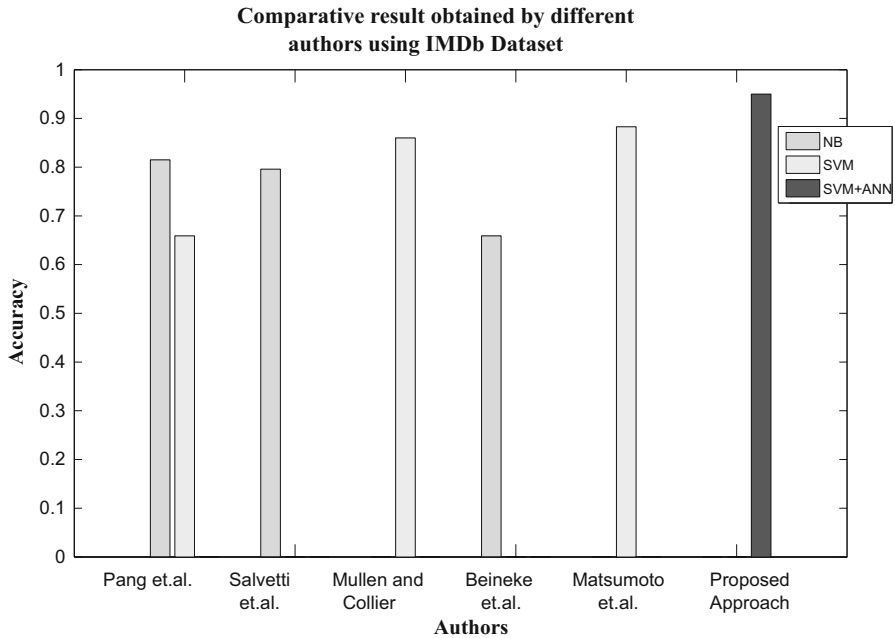


Fig. 4 Comparative result obtained by different literature using IMDb dataset

Table 9 Comparative result obtained by different literature using polarity dataset

Authors	Machine learning techniques used	Accuracy
Pang and Lee [2]	Naive Bayes (NB), SVM	NB = 0.864 , SVM = 0.872
Whitelaw <i>et.al.</i> [41]	SVM	SVM = 0.902
Matsumoto <i>et.al.</i> [11]	SVM	SVM = 0.937
Aue and Gamon [42]	SVM	SVM = 0.905
Read [43]	NB, SVM	NB = 0.789 , SVM = 0.815
Kennedy and Inkpen [44]	SVM	SVM = 0.862
Moraes <i>et.al.</i> [12]	NB, SVM,	NB = 0.803, SVM = 0.841, ANN = 0.865
Proposed approach	Hybrid of SVM and ANN	SVM + ANN = 0.964

5.1 Managerial insights based on result

The managerial insight based on the obtained result can be explained as follows:

- It was almost an observed practice that sellers send questionnaires to the customers, about the feedback of the product they have bought. But nowadays buyers or users share those views through reviews or blogs.
- The reviews can be collected and given input to the proposed approach for qualitative decisions.
- The proposed approach classifies the reviews into either positive or negative polarity and hence is able to guide the managers properly about the shortcoming or good features of the product for future decision making.

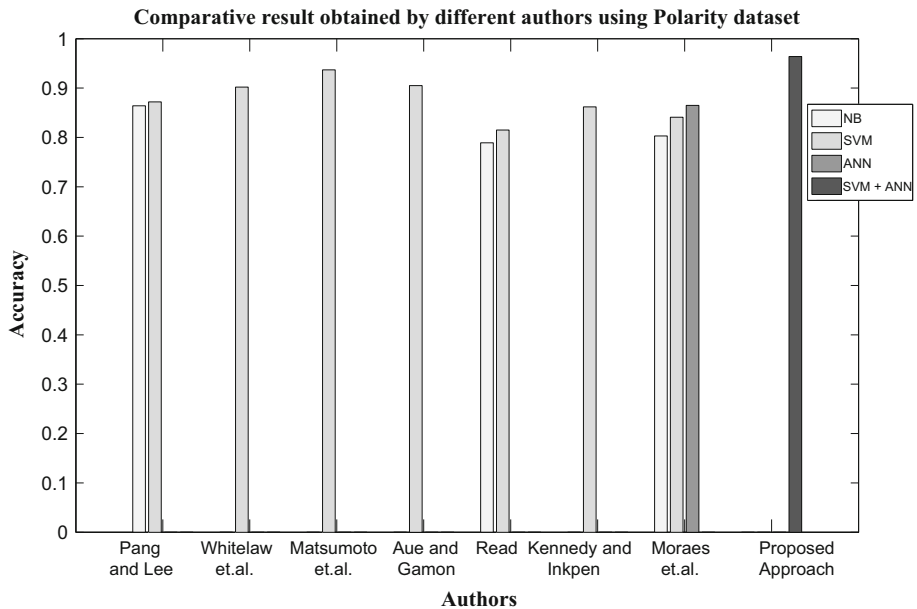


Fig. 5 Comparative result obtained by different literature using polarity dataset

- Every product needs improvement in course of time, and also updates are needed for the product to be useful in all circumstances. But for that reason, the comments about the product must be obtained from the users.
- The users share their comments after use. The comments are collected and as they are not labeled one or if labeled mostly in five-star rating. The reviews are analyzed properly that suggest the issue related to the product to keep it up to date.
- By collecting the reviews about the product with in a specific time interval, analyzing it and then implementing them keep the quality of the product which help the product to be successful in the competition market.

6 Conclusion and future work

In this present study, an attempt has been made to combine two different machine learning algorithms, i.e., SVM and ANN, in order to classify the sentiments of review data associated with movies. As the SVM method analyzes most of the features, i.e., words present in the review, it provides a sentiment value to it. Then using the sentiment value as criteria, it selects the best features from the list of features. By varying the no. of neurons in the hidden layer of ANN, the value of accuracy is found out. The final output indicates whether the review is of positive or negative in nature. The accuracy obtained in this method is found to be comparably better than the results obtained by other authors.

The present study has also some limitations as mentioned below:

- The Twitter comments are mostly small in size. Thus, the proposed approach may have some issues while considering these reviews.

- Different reviews or comments contain symbols like (☺, ☹, 👍, 👎) which help in presenting the sentiment, but these images are not taken into consideration in this study for analysis.
- In order to give stress on a word, it is observed that some persons often repeat the last character of the word a number of times such as “greatttt, fineee.” These words do not have a proper meaning, but they may be considered and further processed to identify sentiment. However, this aspect is also not considered in this paper.

In future, all of above-mentioned limitations may be considered for the future work, in order to improve the quality of sentiment classification. Again an attempt may be made to add few linguistic features while performing the classification of reviews.

References

1. 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, vol 10, Association for Computational Linguistics, 2002, pp 79–86
2. 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, 2004, p 271
3. Turney PD (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, 2002, pp 417–424
4. Liu B (2012) Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol* 5(1):1–167
5. Feldman R (2013) Techniques and applications for sentiment analysis. *Commun ACM* 56(4):82–89
6. Gautam G, Yadav D (2014) Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In: 2014 seventh international conference on contemporary computing (IC3), IEEE, 2014, pp 437–442
7. Hastie T, Tibshirani R, Friedman J (2009) *Unsupervised learning*. Springer, Berlin
8. Hady MFA, Schwenker F (2013) Semi-supervised learning. In: Bianchini M, Maggini M, Jain LC (eds) *Handbook on neural information processing*. Springer, Berlin, pp 215–239
9. IMDb, Internet movie database (IMDb) (2011). http://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz
10. Garreta R, Moncecchi G (2013) *Learning scikit-learn: machine Learning in Python*. Packt Publishing Ltd, Birmingham
11. Matsumoto S, Takamura H, Okumura M (2005) Sentiment classification using word sub-sequences and dependency sub-trees. In: Ho TB, Chung D, Liu H (eds) *Advances in knowledge discovery and data mining*. Springer, Berlin, pp 301–311
12. Moraes R, Valiati JF, Neto WPG (2013) Document-level sentiment classification: an empirical comparison between SVM and ANN. *Expert Syst Appl* 40(2):621–633
13. Tang D (2015) Sentiment-specific representation learning for document-level sentiment analysis. In: Proceedings of the eighth ACM international conference on web search and data mining, ACM, 2015, pp 447–452
14. Tu Z, He Y, Foster J, van Genabith J, Liu Q, Lin S (2012) Identifying high-impact sub-structures for convolution kernels in document-level sentiment classification. In: Proceedings of the 50th annual meeting of the Association for Computational Linguistics: short papers, vol 2, Association for Computational Linguistics, 2012, pp 338–343
15. Liu SM, Chen J-H (2015) A multi-label classification based approach for sentiment classification. *Expert Syst Appl* 42(3):1083–1093
16. Zhang D, Xu H, Su Z, Xu Y (2015) Chinese comments sentiment classification based on word2vec and SVM perf. *Expert Syst Appl* 42(4):1857–1863
17. Luo B, Zeng J, Duan J (2016) Emotion space model for classifying opinions in stock message board. *Expert Syst Appl* 44:138–146
18. Niu T, Zhu S, Pang L, El Saddik A (2016) Sentiment analysis on multi-view social data. In: Tian Q, Sebe N, Qi G, Huet B, Hong R, Liu X (eds) *Multimedia modeling*. Springer, Berlin, pp 15–27
19. Tripathy A, Agrawal A, Rath SK (2016) Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst Appl* 57:117–126

20. Govindarajan M (2013) Sentiment analysis of movie reviews using hybrid method of naive bayes and genetic algorithm. *Int J Adv Comput Res* 3(4):139
21. Abbasi A, Chen H, Salem A (2008) Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. *ACM Trans Inf Syst (TOIS)* 26(3):12
22. Balage Filho PP, Avanço L, Pardo TA, Nunes MG (2014) NILC USP: an improved hybrid system for sentiment analysis in Twitter messages. *SemEval 2014*:428
23. Jagtap B, Dhotre V (2014) SVM and HMM based hybrid approach of sentiment analysis for teacher feedback assessment. *Int J Emerg Trends Technol Comput Sci (IJETCS)* 3(3):229–232
24. Wang S, Wei Y, Li D, Zhang W, Li W (2007) A hybrid method of feature selection for Chinese text sentiment classification. In: Fourth international conference on fuzzy systems and knowledge discovery, 2007 (FSKD 2007), vol 3, IEEE, 2007, pp 435–439
25. Babatunde O, Armstrong L, Leng J, Diepeveen D (2014) A genetic algorithm-based feature selection. *Br J Math Comput Sci* 4(21):889–905
26. Neumann J, Schnörr C, Steidl G (2005) Combined SVM-based feature selection and classification. *Mach Learn* 61(1–3):129–150
27. Fernandez-Lozano C, Seoane JA, Gestal M, Gaunt TR, Dorado J, Campbell C (2015) Texture classification using feature selection and kernel-based techniques. *Soft Comput* 19(9):2469–2480
28. Maldonado S, Weber R, Basak J (2011) Simultaneous feature selection and classification using kernel-penalized support vector machines. *Inf Sci* 181(1):115–128
29. Zheng L, Wang H, Gao S (2015) Sentimental feature selection for sentiment analysis of Chinese online reviews. *Int J Mach Learn Cybern* 6:1–10
30. Sharma A, Dey S (2012) A comparative study of feature selection and machine learning techniques for sentiment analysis. In: Proceedings of the 2012 ACM Research in Applied Computation Symposium, ACM, 2012, pp 1–7
31. Hardin D, Tsamardinos I, Aliferis CF (2004) A theoretical characterization of linear svm-based feature selection. In: Proceedings of the twenty-first international conference on machine learning, ACM, 2004, p 48
32. Tang H, Tan S, Cheng X (2009) A survey on sentiment detection of reviews. *Expert Syst Appl* 36(7):10760–10773
33. Refaeilzadeh P, Tang L, Liu H Cross-validation. <http://www.public.asu.edu.tang9/papers/ency-cross-validation.pdf>
34. Hsu CW, Chang CC, Lin CJ (2003) A practical guide to support vector classification. Technical Report, Department of Computer Science, National Taiwan University
35. Zhang GP (2000) Neural networks for classification: a survey. *IEEE Trans Syst Man Cybern C Appl Rev* 30(4):451–462
36. Reby D, Lek S, Dimopoulos I, Joachim J, Lauga J, Aulagnier S (1997) Artificial neural networks as a classification method in the behavioural sciences. *Behav Process* 40(1):35–43
37. Mouthami K, Devi KN, Bhaskaran VM (2013) Sentiment analysis and classification based on textual reviews. In: 2013 international conference on information communication and embedded systems (ICI-CES), IEEE, 2013, pp 271–276
38. Salvetti F, Lewis S, Reichenbach C (2004) Automatic opinion polarity classification of movie. *Colo Res Linguist* 17:2
39. Mullen T, Collier N (2004) Sentiment analysis using support vector machines with diverse information sources. In: Lin D, Wu D (eds) *EMNLP*, vol 4, pp 412–418
40. Beineke P, Hastie T, Vaithyanathan S (2004) The sentimental factor: improving review classification via human-provided information. In: Proceedings of the 42nd annual meeting on Association for Computational Linguistics, 2004, p 263
41. 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, ACM, 2005, pp 625–631
42. Aue A, Gamon M (2005) Customizing sentiment classifiers to new domains: a case study. In: Proceedings of recent advances in natural language processing (RANLP), vol. 1, 2005, pp 1–7
43. Read J (2005) Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL student research workshop, Association for Computational Linguistics, 2005, pp 43–48
44. Kennedy A, Inkpen D (2006) Sentiment classification of movie reviews using contextual valence shifters. *Comput Intell* 22(2):110–125
45. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: European conference on machine learning, pp 137–142

46. Socher R, Perelygin A, Wu JY, Chuang J, Manning C, Ng A, Potts C (2013) Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP), pp 1642–1654
47. Cao Y, Xu R, Chen T (2015) Combining convolutional neural network and support vector machine for sentiment classification. In: Chinese national conference on social media processing, pp 144–155
48. Liu B (2015) Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge University Press, Cambridge
49. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint [arXiv:1301.3781](https://arxiv.org/abs/1301.3781)
50. van Rijsbergen CJ, Robertson SE, Porter MF, Martin F (1980) New models in probabilistic information retrieval. British Library Research and Development Department, London
51. Goldberg Y, Levy O (2014) word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint [arXiv:1402.3722](https://arxiv.org/abs/1402.3722)
52. Blake C, Merz CJ (1998) {UCI} Repository of machine learning databases. University of California, Dept. of Inform. Computer science, Irvine, CA, Available: <http://www.ics.uci.edu/mllearn/ML-Repository.html>
53. Weston J, Elisseeff A, Schölkopf B, Tipping M (2003) Use of the zero-norm with linear models and kernel methods. *J Mach Learn Res* 3:1439–1461



Abinash Tripathy is currently working as Assistant Professor at National Institute of Science and Technology, Berhampur, Odisha, India. He obtained his Ph.D. from National Institute of Technology, Rourkela, Odisha. His research interests are software testing, UML, natural language processing and sentiment analysis.



Abhishek Anand is currently working as Software Engineer at Microsoft India Private Limited. He obtained his B. Tech. from National Institute of Technology, Rourkela, Odisha, in Computer Science and Engineering. His research interests are software testing, software engineering, natural language processing and sentiment analysis.



Santanu Kumar Rath is a Professor in the Department of Computer Science and Engineering, National Institute of Technology, Rourkela, since 1988. His research interests are software engineering, system engineering, bioinformatics, natural language processing and management. He is a Senior Member of the IEEE, USA, and ACM, USA, and Petri Net Society, Germany.