International Journal of Latest Trends in Engineering and Technology
Vol.(8)Issue(4-1), pp.020-031
DOI: http://dx.doi.org/10.21172/1.841.04
e-ISSN:2278-621X

COMPARATIVE STUDY OF DEEP LEARNING BASED SENTIMENTAL ANALYSIS WITH OTHER EXISTENCE TECHNIQUES

Seema Choudhary[1](Research Scholar), Sandeep Bhargava[2]

Abstract-In today's era, social networking sites have become pervasive and a basic component, they have emerged as a major platform for people to express their feelings. The essence of the social content posted, the exact feelings hidden in the exposed content can be greatly analyzed by sentimental analysis. This survey paper presents a "Comparative study of deep learning based sentiment analysis with other existence techniques and challenges". Deep learning techniques have shown impressive performance across various natural language processing tasks and multiple data sets. In this review, we are going to explore a comparison review between various algorithms and challenges in the area of sentiment analysis. Also this paper highlights applications of sentiment analysis or opinion mining.
Keywords – Sentiment analysis, Machine learning techniques, Deep learning techniques, Lexicon-based techniques, Hybrid techniques

1. INTRODUCTION

By the virtue of its nature, sentiment analysis refers to the utilization of language process or text analysis with the help of which emotive states and subjective data can be easily established, extracted and quantified. A part from this, it also encounters an emotional reaction and interaction with a particular document or event. Such technique is also accustomed to verify sentiments on different levels. It helps in aggregating the complete document either positive or negative. As well as, it helps in formulating the combined reaction of individual words or phrases present in the document. It basically tracks a given selected topic. Various companies use such technique in analyzing their services, merchandise. As an instance, if somebody becomes offensive as a whole on social media, sentiment analysis grades such activity as negative and thus alerts for such activities can be yielded with hyper-negative sentiment scores.
The rest of the paper is organized as follows: Section II describes the applications of sentiment analysis; Section III Section discusses the techniques of sentiment analysis; IV describes the comparative review; Further, Section V explains the Key challenges in sentiment analysis. Finally, Section VI concludes the paper.

2. APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment analysis has shown its applicability in various areas which can be described as follows:

2.1 Product and Repair Reviews –

The foremost and common application of sentiment analysis is within the space of reviews of client merchandise and their cumulative services. In this context, such technique amplifies several websites, thus giving automatic summaries of reviews concerning the merchandise and their specific aspects. This can be very helpful in revamping the standards of the product or the service according to the reviews and thus, improving the customer traffic. A notable example can be "Google Product Search".

2.2 Result prediction –

By analyzing opinions or sentiments from relevant sources, one can easily predict the probable outcome of a specific event. As an example, sentiment analysis dispenses the substantial price to candidates running for varied positions. It permits campaign managers to analyze the reaction of voters concerning different problems and the way these voters relates with the actions of candidates.

2.3 Creating, deciding higher cognitive process –

It has also proved its ability in choice-making systems. An example of such approach can be found with the money market investments. Sentiment analysis system uses these sources in order to search the articles that debates the businesses and mixes the sentiment that concerns them as an aggregated score which can further be employed by any automatic commercial system.

[1] Department of Computer science Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India
[2] Department of Computer science Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

3. TECHNIQUES OF SENTIMENT ANALYSIS

Sentiment techniques may be classified into three different approaches which have been described as follows (Fig. 1.):

Machine Learning Approach
Deep learning Approach
Lexicon primarily Based Approach
Hybrid Approach.

The Machine Learning Approach (ML) applies the celebrated cubic centimeter algorithms and uses linguistic options. Deep learning is a subset of machine learning, which employs deep neural networks to learn good representations of the input data, which helps to perform specific tasks. The Lexicon-based Approach depends on a sentiment lexicon which is a set of famous and pre-compiled sentiment terms. Such lexicon based approach is further divided into two approaches which are: dictionary-based approach and corpus-based approach. These approaches use applied mathematics or linguistics strategies to seek out sentiment polarity. The Hybrid Approach combines the above mentioned approaches and is incredibly common with sentiment lexicons taking part in a key role within the majority of strategies.

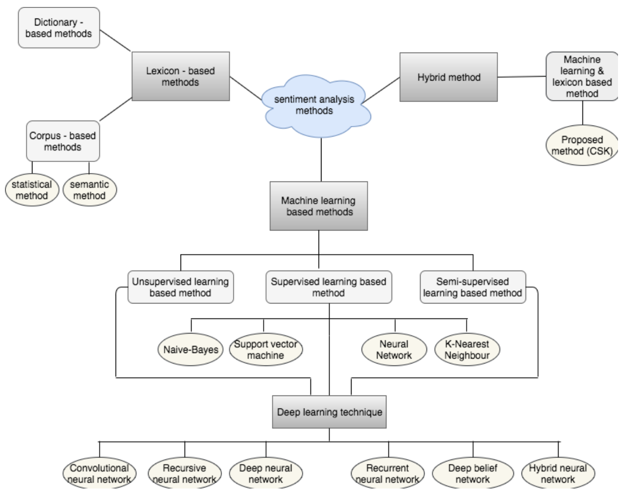


Fig 1Techniques Of Sentiment Analysis

3.1. Machine Learning based Technique –

Machine Learning primarily based sentiment analysis or classification comprises of the following techniques:
Sentiment Analysis by exploiting Supervised Machine Learning Technique and
Sentiment Analysis by exploiting Unsupervised Machine Learning Technique

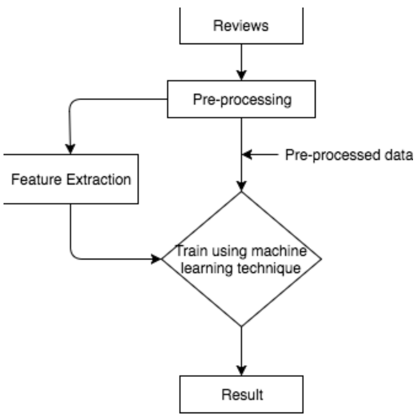


Fig 2 Working Of A Machine Learning Technique

In Supervised Machine Learning technique, there are two types of information that is required: Training Information Set and Test Data Set. An automatic classifier learns the classification factors of the document from the training set and thus the accuracy in the classification is evaluated using the test data set. The test data set plays an important role in validating the classification. In this context, various machine learning algorithms have been proposed in literatures that are used to classify the documents. These machine learning algorithms such as Support Vector Machine (SVM), Naive Bayes (NB) and neural networking are used with success in various analyses and shows its accuracy in sentiment classification.

Naïve Baye-It is one amongst the most effective, widely used and easy approaches for text classification. This approachis basically based on bayes probability theorem. In this approach, the posterior probability of class or given predicator is calculated and the final probability is calculated by multiplying the preceding probability with the likelihood. the strategy is Naïve within the sense that it assumes each word within the text to be independent. This assumption makes it easier to implement however less correct.

P(c/x) = [p(x/c)p(c)] / p(x)
– p(c/x) is the posterior probability of class (target) given predicator (attribute).
– p(c) is the prior probability of class.
– p(x/c) is the likelihood which is the probability predictor given class.
– p(x) is the prior probability of predictor.

Support Vector Machines (SVM)-It is additionally used for text classification based on a discriminative classifier. The approach is predicated on the principle of structural risk reduction. 1st the training information points are separated into 2 totally different categories based on a decided decision criteria or surface. the decision relies on the support vectors selected within the training set. Among the various variants of SVM, the multi category SVM is used for sentiment analysis. The centroid classification algorithmic rule initially calculates the centroid vector for each training class then the similarities between a document and all the centroids are calculated also, the document is allotted a category based on these similarity values.

The K-Nearest Neighbor (KNN) – This approach finds the K nearest neighbors of a text document among the coaching documents. The classification is completed on the idea of the similarity score of the category to the neighbor document. Winnow is another prominent approach that is widely in use. The system initially predicts a category for a specific document then receives the feedback. In presence of false classification which is primarily an error the system, updates its weight vectors accordingly. This method is repeated over a sufficiently massive set of training information.

Neural Network – Basically, Neural network is a part of deep learning. It has many benefits .However one amongst the most recognized of these is the fact that it can really learn from perceptive information sets. During this approach, neural network is used as a random operate approximation tool. This type of tools facilitates to estimate the most cost-efficient and ideal strategies for arriving at solutions while defining computing functions or distributions. A neural network takes information samples instead of entire information sets to make solutions, that saves both time and cash. A neural networks are thought of fairly easy mathematical models to boost existing information analysis technologies. Neural networks have 3 layers that are interconnected. the primary layer consists of input neurons. Those neurons send information on to the second layer, that successively sends the output neurons to the third layer.
The first step in Supervised Machine Learning Technique is to gather the coaching set and to choose the suitable classifier. Once the classifier is chosen, the classifier gets trained thereby exploiting the collected coaching set. The key step within the Supervised Machine Learning Technique is feature choice. The classifier choice and feature choice determines the performance of classification. Several common techniques used for feature choice have been presented as follows:

Opinion words and phrase – Opinion words can be extracted from the document with the help of adjectives and adverbs that are used in the document. Additionally, nouns or verbs can also be used for specifying opinion. For example, good, fantastic, amazing, dangerous and boring can be classified as adjectives or adverbs that categorize emotions whereas rubbish can be classified as a noun since it specifies a sentiment equally hate. Once opinions are collected, their polarity is calculated based on statistical or lexicon techniques. Hu and Liu et al. [4] uses a WordNet Api for decisive polarity orientation for chosen sentiment words.

Terms and their frequency – Uni-grams or n-grams with their frequency of prevalence can be thought of as options. Such method has been employed in several studies and has achieved sensible result. Pang et al. [1] used uni-grams on picture review dataset and Dave et al. [3] used bigrams and trigrams on product review dataset where each studies rumored higher result on polarity determination.

Part of speech (POS) information – In this approach, POS tag of words has been employed in classifying the feature. In POS tagging, every word is tagged by considering its position within the grammatical context. Prabowo and Thelwall [8] has used such approach in their studies and have created feature set simply by the utilization of characteristic adjectives and adverbs.

Negations – Negation word reverses the meaning of the sentence. Therefore, it is important to consider negation in polarity calculation.

3.1.1 Deep Learning –
Deep learning refers to artificial neural networks that are composed of the many layers. It's a growing trend in machine learning because of some favorable results in applications wherever the target operate is extremely complex and also the datasets are massive.

How does it work?
Basically Deep Learning involves feeding computing system lots of information, that it can use to create decisions regarding different information. This data is fed through neural networks, as is that the case in machine learning. These networks – logical constructions that ask a series of binary true/false queries, or extract a numerical worth, of each little bit of information that go through them, and classify it according to the answers received as a result of Deep Learning work is targeted on developing these networks, they become what are called Deep Neural Networks – logic networks of the complexity required to deal with classifying datasets as massive, as say, Google's image library, or Twitter's firehouse of tweets.

Convolutional neural networks – Convolutional neural networks work like learnable native filters. deep convolution neural networks have performed with new achievements in the field of image classification and face and location recognition, for example. These networks use several, extended layers of neurons to construct autonomously more and more abstract, very native and elaborated representations of a picture. The best example is maybe their application to computing vision. the primary step in image analysis is usually to perform some native filtering of the image, for instance, to boost edges within the image. you are doing this by taking the neighborhood of every picture element and flex it with an explicit mask. essentially you reckon a linear combination of these pixels. for instance, if you have a positive weight on the middle pixel and negative weights on the surrounding pixels you compute the difference between the center pixel and the surrounding, giving you a crude kind of edge detector. Now you can either put that filter in there by hand or learn the correct filter through a convolutional neural network. If we tend to consider the easy case, you have an input layer representing all pixels in your image whereas the output layer representing the filter responses. every node in the output layer is connected to a pixel and its neighborhood in the input layer. So far, so good. What makes CNN special is that the weights are shared, that is, they're a similar for various pixels in the image (but different with respect to the position relative to the middle pixel). In this method you effectively learn a filter, which turns out to be suitable to the problem you're attempting to learn.

Recursive neural network – A recursive neural network is additional like a hierarchical network where there's really no time aspect to the input sequence but the input has to be processed hierarchically in a tree fashion. recursive neural networks are very powerful just because they represent each word as a vector and an operator that to me seems very intuitive. so the word "not" can be a rotation matrix that acts on subsequent word (for eg. fine) and changes it's polarity by rotating the vector of fine to currently mean not good. this is a very powerful thought. however these networks need lots of training knowledge (parse trees are needed to train these networks). I must re-iterate once more that this kind of representation is very intuitive (although easier neural network techniques work better).

Deep-neural networks – A deep-neural network usually has a definite structure. This structure provides lots towards the performance of the neural network. Usually, there's an initial layer, a hidden layer and a final layer. This architecture was acceptable for finding a number of issues. but the error rate was still quite high. Thus, a deep architecture of neural networks was developed. It has an input layer, several hidden layers and an output layer. This architecture was developed to enhance accuracy, but at the price of potency and its application wasn't possible till modern-day GPUs came on to enhance efficiency. The accuracy of the neural networks were found to extend as the number of hidden layers raised. In other words, as the neural network got "deeper" in terms of architecture, it performed best and best. Although, that's only one factor that improves accuracy. other factors include, little averaging pools and strides etc.

Recurrent neural network – A repeated neural network essentially unfolds over time. it's used for ordered inputs where the time factor is the main differentiating factor between the elements of the sequence. recurrent neural networks are neural networks during which nodes can be connected to both downstream and upstream nodes. by contrast feedforward networks (which are usually organized in "layers") only permit upstream nodes to directly influence downstream nodes. Activation flows around recurrent networks instead of simply through them from input to output. not like feed-forward networks RNNs can maintain internal state. Whereas it makes sense to consider the weights during a feedforward network implementing a mapping from input to output, it makes sense to consider the weights during a recurrent network shaping the dynamics of the network activation flow. A subset of RNNs are CTRNNs - continuous time recurrent neural networks that update the network in continuous time instead of separate phases, and can be used as control systems for easy mobile robots.

Deep belief network – Deep belief networks are probabilistic generative methods that are composed of more than one layers of stochastic, latent variables. The latent variables typically have binary values and are usually called hidden units or feature detectors. the highest 2 layers have directionless, symmetrical connections between them and type of an associative memory. The lower layers receive top-down, directed connections from the layer above. The states of the units in the lowest layer represent an information vector.

The two most vital properties of deep belief networks are –

There is an efficient, level-by-level process for learning the top-bottom, generative weights that verify how the variables in one layer rely on the variables in the layer above.

After learning, the values of the latent variables in each layer can be inferred by one, bottom-up pass that starts with a discovered information vector in the bottom layer and uses the generative weights in the opposite direction.

DBN are learned one layer at a time by treating the values of the latent variables in one layer, when they are being inferred from information, as the information for training subsequent layer. This efficient, greedy learning can be followed by, or combined with, other learning procedures that fine-tune all of the weights to enhance the generative or discriminative performance of the entire network. Discriminative fine-tuning can be presented by adding a output layer of variables that represent the required outputs and back propagating fault derivatives. when networks with several hidden layers are applied to highly-structured input file, like pictures, back propagation works far better if the feature detectors within the hidden layers are initialized by learning a deep belief internet that models the structure in the input file.

Hybrid neural network – These type of methods are usually carried with the combination of symbolic computation and deep neural networks into one model Symbolic representations have benefits with relevancy specific, direct management, quick initial committal to writing, dynamic variable binding and data abstraction. Representations of deep neural networks, on the opposite hand, show benefits for biological credibility, learning, vitality (fault-tolerant process and swish decay), and generalization to similar input.

3.2 Lexicon Based Method –

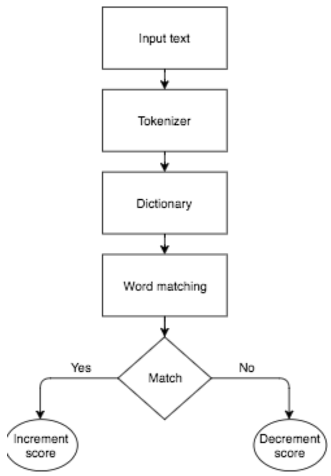
Lexicon primarily based technique is an unsupervised Learning approach since it does not need any prior coaching information sets. Such approach can be categorized as linguistics orientation approach for opinion mining. In this approach, sentiment polarity of options specified in the given document are determined by comparing the mentioned options with linguistic lexicons. These linguistic lexicons comprises of a lost of words whose sentiment oriented is already determined. This approach classifies the document by aggregating the sentiment orientation of all opinion words present within the given document. In result, documents with additional positive word lexicons can be classified as positive document, whereas, the documents with additional negative word lexicons can be classified as negative document. The key steps used in the process of lexicon primarily based sentiment analysis are listed as follows:

Preprocessing – In preprocessing step, unnecessary HTML tags and screaming characters are eradicated from the document. This step is also responsible for correcting writing system mistakes, descriptive linguistic mistakes, punctuation errors and incorrect capitalization. Apart from this, utilization of actual terms is made for non-dictionary words.

Feature Selection – This step is responsible for extracting the features present within the document with the help of exploitation techniques such as POS tagging.

Sentiment Score Calculation – For this step, let us consider a score say ‘s’ and initialize it with zero. Now, for every extracted sentiment word, check whether it is present within the sentiment lexicon or not. If it is present with negative polarity say ‘w’, then s = s - w and if it is present with positive polarity, then s = s + w.

Sentiment Classification – If the score ‘s’ calculated in the previous step is below a specific threshold worth, then document is classified as negative otherwise it is classified as positive.



Working of a Lexical Technique
Sentiment lexicon can be constructed in three ways which has been illustrated as follows:

Manual Lexicon Construction – Such type of construction facilitates the manual construction of lexicons. It is usually time consuming and is terribly tough.

Dictionary-Based Lexicon Construction – In this type of lexicon construction, a small set of sentiment words and their polarity is calculated manually and is then widened by the addition of words thereby making the exploitation WordNet wordbook or SentiWordNet wordbook with their corresponding synonyms and antonyms.

Corpus-Based Lexicon Construction – This type of lexicon construction works by considering grammar patterns of the words that are present within the document. In order to produce correct linguistic words, annotated coaching knowledge is required

3.3 Hybrid Techniques –
These type of techniques are usually carried with the combination of supervised machine learning approaches and lexicon primarily based approaches. Such techniques facilitate the enhancement of sentiment classification performance. In this context, Fang et al. [18] adopted an entirely different approach. The researchers considered every general purpose lexicon and domain specific lexicon for decisive polarity orientation of sentiment words which are then reinforced to supervised learning rule, SVM. With their analysis, they concluded that the performance of domain specific lexicon is better than the performance of general purpose lexicon. The system classified the sentiments in two steps: initially, the classifier is trained to predict the aspects and then the classifier is trained to predict the feelings associated with the aspects collected in the previous step. The applied methodology provided an accuracy of 66.8%.
Mudinas et al. [24] combined lexicon primarily based and learning based approaches for developing a concept-level sentiment analysis system called pSenti. With linguistic lexicon methodology, stability and readability was attained, whereas, accuracy was attained from supervised learning rule. In such approach, initially sentiment words were extracted which were then considered as options. This hybrid approach pSenti attained an overall accuracy of 82.30%.
Zhang et al. [19] distributed entity level sentiment analysis. For such approach, researchers used supervised learning techniques and lexicon primarily based techniques. Sentiments words were extracted with the usage of lexicon primarily based technique. Further, seeds were discovered by implementing exploitation Chi-square check on the extracted seeds so far. As soon as the seeds are discovered, sentiment polarity of recently discovered seed is determined through a classifier that is already being trained exploitation initial seeds. Such methodology achieved an accuracy of 85.4% and requires no manual effort.

4. COMPARATIVE REVIEW

This section gives a detail on the comprehensive review of 25 sentimental analysis research papers. The exposition of various articles/papers have been outlined in Table 1.1.

S. No	Year	Reference	Research – Work	Algorithm	Data - set	Challenge
1	2010	Bing liu [11]	opinion analysis	Machine learning	Product - review	Sentiment polarity categorization
2	2010	Fangtao et al. [14]	Investigation of the sentiment dependency in joint sentiment and topic analysis	Sentiment LDA and Dependency sentiment LDA	Customers review	Domain dependence
3	2011	Bas et al. [15]	Investigate of the impact of accounting for negation in sentiment analysis	Part of speech (POS)	Dutch language	Negation
4	2011	Yulan et al. [16]	polarity-bearing topics generated from the JST model	Naive Bayes and support vector machine from WEKA5	Movie reviews	Domain dependence
5	2011	Lei Zhang et. al. [19]	A new entity - level sentiment analysis method for twitter	Lexicon - based & Machine learning	Twitter	Low performance of recall

S. No	Year	Reference	Research – Work	Algorithm	Data - set	Challenge
6	2011	Alexander et al. [20]	optimize the sentiment modification in case of negation to a value of –1.27 rather than –1	Part of speech (POS)	Movie reviews	Negation
7	2011	Jiang and Min [21]	sentiment sentence based approach to explore the overall sentiment polarity of Chinese reviews	n - gram	Chinese reviews	Domain - dependence
8	2011	Walter and Mihaela [22]	web based opinion mining system for hotel reviews	n - gram	Hotel reviews	Domain - dependence and Features Extraction
9	2011	Myle et al. [23]	developed the first large-scale dataset containing gold-standard deceptive opinion spam	POS and n - gram	Customers reviews	Spam or fake reviews
10	2012	Andrius et al. [27]	A concept-level sentiment analysis system that seamlessly integrates into opinion mining lexicon-based and learning-based Approaches	Support vector Machine approach	Software and movie reviews	Huge lexicon
11	2013	Alexandra et al. [30]	summarized sentiment classification for news and applied different methods to test the appropriateness of different resources and approaches to the task defined	Lexicon - based approach	Newspaper articles	Domain - dependence
12	2013	Ivan et al. [32]	in-depth research of supervised machine learning methods for sentiment analysis of Czech social media	n - gram	Facebook	Bi-polar words
13	2013	Hemalatha et. al. [33]	Twitter sentiment analysis	Machine learning approach	Twitter	Domain - dependence
14	2014	Pappu Rajan & S. P. Victor [34]	Web sentiment analysis	Lexical approach	Twitter	Sentiment polarity categorization
15	2014	Qingxi Peng & Ming Zhong [35]	Detecting spam review	Combine lexicon and use shallow dependency parser	Store#364	Spam and fake reviews
16	2015	Rajni singh & Rajdeep Kaur [45]	Develops a combined dictionary based on social media keywords & online review	Open source Data mining tool	Online review & Twitter	Domain - dependence
17	2015	Xing Fang & Justin Zhan [46]	Tackle the problem of sentiment polarity categorization	Naïve Bayesian, Random Forest, and Support Vector Machine.	Amazon	sentiment polarity categorization

S. No	Year	Reference	Research – Work	Algorithm	Data - set	Challenge
18	2015	Lucie Flekova et. al. [47]	Analyzing domain suitability of a sentiment analysis	n - gram (uni & bi-grams)	Crowdsourced annotations & Facebook	bipolar words
19	2015	V. Pream Sudha et. al. [48]	Review on deep learning	Neural network	Google now and apple'siri	Huge lexicon
20	2016	Neeru Mago [50]	Opinion mining	Machine learning approach	Web - based	domain dependency
21	2016	Salma Farooq [56]	Survey on Opinion Spam Detection	Machine learning technique	Online reviews	Spam and fake reviews
22	2016	Lina Maria Rojas-Barahona [57]	An overview of deep learning for sentiment analysis	Recursive and non recursive neural network	Movie-review, twitter and sentiment treebank	Sentiment polarity categorization
23	2016	Yuhai Yu et. al. [58]	Analyzing visual and textual sentiments using by deep convolutional neural networks.	Deep convolutional neural network	Twitter and Sina Weibo	Huge lexicon and feature extraction
24	2017	Hajra Wahed et. al. [59]	Investigation of sentiment analysis	Thematic analysis method	Social-sites	Sentiment polarity categorization
25	2017	Qurat Tul Ain et. al. [60]	survey on Sentiment Analysis Using Deep Learning Techniques	Deep-learning technique	Social-sites	Sentiment polarity categorization

Table 1.1 *Comparison of various machine-learning algorithms –*

Algorithm	Features	Drawbacks	Benefits	Efficiency
Naive - Bayes Algorithm	• Widely used and a easy approach for text classification. • Require a small amount of trading data to estimate the parameter.	• The precision of algorithm decreases if the amount of data is less. • For obtaining good results it requires a very large number of records.	• Simple to implement. • It predicts accurate results for most of the classification and prediction problems.	• Most effective but Less accuracy.

Support vector machine (SVM) algorithm	It is also used for text classification based on a discriminative classifier.	• Speed and size requirements both in training and testing is more. • High complexity and extensive memory requirements for classification in memory cases.	• Work well even if data is not linearly separable in the base feature space.	• High accuracy.
Neural - Network algorithm	• Neural network has emerged as an important tool for classification. • During past decade neural network classification has established as a promising alternative to various conventional classification methods.	• Require high processing time if neural network is large. • Needs massive data set and hugely expensive to train.	• Easy to implement. • Application to a problems in real life. • It is easy to use, with few parameters to adjust. • A neural network learns and reprogramming is not needed. • performing in supervised, semi-supervised and unsupervised types.	• Great accuracy compare to SVM .
K - Nearest neighbor algorithm	• This approach finds the K nearest neighbors of a text document among the training documents.	• Time to find the nearest neighbors in a large training data set can be excessive. • It is sensitive to noisy or irrelevant attributes. • Performance of algorithm depends on the number of dimensions used.	• Classes need not be linearly separable. • Sometimes it is robust with regard to noisy training data. • Well suited for multi - modal classes.	• Less cost efficiency.

5. KEY CHALLENGES IN SENTIMENT ANALYSIS

This survey discusses various challenges that are faced in analyzing the sentiments. From the twenty five papers that have been reviewed , the challenges can be elucidated as :

5.1 Domain Dependency – There are many words whose polarity changes from domain to domain. Taking the below mentioned subsequent examples:
- The movie was unpredictable.
- The steering of the car is unpredictable.

In the first sentence, the sentiment sent is positive whereas the sentiment in second sentence is negative.[14, 16, 21, 22, 30, 33, 50, 56].

5.2 Sentiment polarity categorization – The most challenge concerns with the sentiment categorizing the polarity of a given text at the document, sentence, or feature/aspect level—whether the feelings expressed on SNS are usually categorized as positive, negative or neutral. but a post will contain parts expressing bipolar sentiments or opinions, a feature that should be tackled.

The main reason for this problem is the natural language. In a text document same word may have multiple meanings and same phrase or sentence can be interpreted in different ways which leaves us a challenge in determining the polarity of a single sentence. [11, 37, 51, 63, 66, 67].

5.3 Negation – Negation in sentiment analysis is that the task of turning positive into negative or vice-versa. There are various types of negations which can be direct (or) indirect, the sentences with these negations (both direct and indirect) act as impediments in choosing the polarity of the sentiment.

Example –

Direct:- I don't like traveling.

Indirect:- I actually love chocolates, but I don't feel the same that. [15, 20].

5.4 Low recall performance – Recall measures the completeness, or sensitivity of a classifier. Higher recall means that less false negatives, whereas Lower recall means that more false negatives rising recall will usually decrease preciseness because it gets progressively harder to be precise as the sample space increases. [19].

5.5 Features Extraction – Feature extraction is a domain-dependent task. during this challenge we tend to produce a feature based opinion summary of multiple reviews.
determine and extract object features that have been commented as a review on by opinion holder. (eg. “room”, “service”, etc.)
Analysis polarity of opinion on features. Like- positive, negative, and neutral. [22, 64].

5.6 Spam and Fake review – During the web review spam or fake reviews that attempt to mislead viewers or customers by giving underserving positive opinions to some target entities in order to promote the entities and by giving in order to harm their reputations. in this case, viewers and customers can’t be able to confirm that which reviews or comments are real or fake. [23, 38, 62].
Example –
App Review for Windows Phone version not iPhone version iOS does not have live tiles, windows phone does. Most likely the developer hired a firm to post fake reviews for both platforms and they got confused.

5.7 Huge Lexicon – Sentiment analysis using a large amount of lexicons or informations during analyzing of people’s opinion. Billion of people express their feelings, opinion concerning various services or products using SNS, blogs, and review sites. during huge lexicon dataset, sentiment analysis faces complexity in categorization. [27, 53, 64].

5.8 Bi-polar words – Those words which represents both type of category like positive or negative sentiments in same sentence at same time. [32, 52].
Example –
“At night, I can’t sleep,
In morning, I can’t wake up.”

6. CONCLUSION

After reviewing available research papers, we may have reached the conclusion that the Neural network, Navie-Bayes, and SVM are the most appropriate machine learning techniques for solving sentiment classification error. With SVM, time value can be reduced and better accuracy and performance can be achieved. the lexicon-based technique used is wordNet that's needed for sentiment analysis tasks. Such technique achieves less accuracy and fewer categorization drawback.
Consistent with review, Deep learning techniques are better than SVMs and other machine learning strategies as a result of the deep learning have more hidden layers as compared to normal networks. These techniques are capable of performing in supervised, semi-supervised and unsupervised form. It can be successfully applied to massive or large knowledge for knowledge discovery, data application, and knowledge-based prediction. In other words, Deep learning can be a robust engine for producing actionable results. this technique also has some drawbacks like needs massive data set and hugely expensive to train.

7. REFERENCES

[1] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp. 79-86.

[2] P. Turney, “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews”, Proceedings of the Association for Computational Linguistics (ACL), 2002, pp. 417– 424.

[3] K. Dave, S. Lawrence, and D. M. Pennock,“Mining the peanut gallery: Opinion extraction and semantic classification of product reviews,” Proceedings of WWW, 2003, pp. 519–528.

[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," Proceedings of the tenth ACM international conference on Knowledge discovery and data mining, Seattle, 2004, pp. 168-177.

[5] A. Abbasi, H. Chen, and A. Salem, “Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums,” In ACM Transactions on Information Systems, vol. 26 Issue 3, pp. 1-34, 2008.

[6] A. Harb, M. Planti, G. Dray, M. Roche, Fran, o. Trousset and P. Poncelet, "Web opinion mining: how to extract opinions from blogs?", presented at the Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, Cergy-Pontoise, France, 2008.

[7] M. Annett, G. Kondrak, “A comparison of sentiment analysis techniques: Polarizing movie Blogs”, In Canadian Conference on AI, pp. 25–35,2008.

[8] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach" , Journal of Informetrics, vol. 3, pp.143-157, 2009.

[9] T. Peng, C. Shih, “An Unsupervised Snippet-Based Sentiment Classification Method for Chinese Unknown Phrases without Using Reference Word Pairs.” Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp.243-248.

[10] Saif, M.M., Peter, D.T., 2010. “Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon”. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Association for Computational Linguistics, Los Angeles, California.

[11] Bing Liu, “Sentiment Analysis and Subjectivity”. In Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010.

[12] Ahmed, A., Zhu, Z., David, Z., Hsinchun, C., Jay, F.N., 2010. “Detecting fake websites: the contribution of statistical learning theory”. MIS Q. 34 (3), 435–461.

[13] Ainur, Y., Yisong, Y., Claire, C., 2010. “Multi-level structured models for document-level sentiment classification”. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. MIT, Massachusetts, Association for Computational Linguistics, USA, pp. 1046–1056.

[14] Fangtao, L., Minlie, H., Xiaoyan, Z., 2010. “Sentiment analysis with global topics and local dependency”. In: AAAI Conference on Artificial Intelligence.

[15] Bas, H., Paul, V.I., Alexander, H., Flavius, F., Uzay, K., 2011. “Accounting for Negation in Sentiment Analysis” DIR 2011. The Netherlands, Amsterdam.

[16] Yulan, H., Chenghua, L., Harith, A., 2011. “Automatically extracting polarity-bearing topics for cross-domain sentiment classification”. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Portland, Oregon.

[17] A. Khan, B. Baharudin, K. Khan; “Sentiment Classification from Online Customer Reviews Using Lexical Contextual Sentence Structure” ICSECS 2011: 2nd International Conference on Software Engineering and Computer Systems, Springer, pp.317-331, 2011.

[18] Ji Fang and Bi Chen, “Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classication”, In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), pages 94–100, 2011.

[19] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B.Liu, “Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis”, Technical report, HP Laboratories, 2011.

[20] Alexander, H., Paul, V.I., Bas, H., Flavius, F., Uzay, K., 2011. “Determining negation scope and strength in sentiment analysis”. In: Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011). IEEE Computer Society.

[21] Jiang, Y., Min, H., 2011. “Using topic sentiment sentences to recognize sentiment polarity in Chinese reviews”. In: CIPS-SIGHAN Joint Conference on Chinese Language.

[22] Walter, K., Mihaela, V., 2011. “Sentiment analysis for hotel reviews”. In: Proceedings of the Computational Linguistics-Applications Conference, JACHRANKA.

[23] Myle, O., Yejin, C., Claire, C., Jeffrey, T., 2011. “Finding deceptive opinion spam by any stretch of the imagination”. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pp. 309–319, Portland, Oregon.

[24] A. Mudinas, D. Zhang, M. Levene, “Combining lexicon and learning based approaches for concept level sentiment analysis”, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York,NY, USA, Article 5, pp. 1-8, 2012.

[25] W. Zhang, H. Xu, W. Wan, “Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis,” Expert Systems with Applications, Elsevier, vol. 39, 2012, pp. 10283-10291.

[26] Amna, A., 2012. “Negation identification and calculation in sentiment analysis”. In: (IMMM 2012) The Second International Conference on Advances in Information Mining and Management.

[27] Andrius, M., Dell, Z., Mark, L., 2012. “Combining lexicon and learning based approaches for concept-level sentiment analysis”. In: WIS-DOM’12, Beijing, China.

[28] Arjun, M., Bing, L., Natalie, G., 2012. “Spotting fake reviewer groups in consumer reviews”. In: International World Wide Web Conference (WWW-2012), Lyon, France.

[29] Arjun, M., Vivek, V., Bing, L., Natalie, G., 2013. “What yelp fake review filter might be doing”. In: Proceedings of The International AAAI Conference on Weblogs and Social Media (ICWSM-2013), Boston, USA.

[30] Alexandra, B., Ralf, S., Mijail, K., Vanni, Z., Erik, V.D.G., Matina, H., Bruno, P., Jenya, B., 2013. “Sentiment analysis in the news”. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10).

[31] Christine, L., Florian, K., Antal, V.D.B., 2013. “The perfect solution for detecting sarcasm in tweets”. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Association for Computational Linguistics, Atlanta, Georgia, pp. 29–37.

[32] Ivan, H., Tomas, P., Josef, S., 2013. “Sentiment analysis in Czech social media using supervised machine learning”. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Association for Computational Linguistics, Atlanta, Georgia, pp. 65–74.

[33] Hemalatha, Dr. G. P Saradhi Varma, Dr. A.Govardhan, “Sentiment Analysis Tool using Machine learning Algorithms”. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 2, Issue 2, March – April 2013.

[34] Basant Agarwal, Namita Mittal, “Optimal Feature Selection for Sentiment Analysis”, In 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013),Vol-7817, pages-13-24, Greece, Samos. 2013.

[35] Basant Agarwal, Vijay Sharma, Namita Mittal, “Sentiment Classification of Review Documents using Phrases Patterns”, In Second International Symposium on Natural Language Processing (NLP’13). pp.1577-1580, 2013.

[36] Basant Agarwal, Namita Mittal, Erik Cambria “Enhancing Sentiment Classification Performance using Bi-Tagged Phrases”, In SENTIRE , ICDM Workshop. pp: 892-895 IEEE. 2013.

[37] A Pappu Rajan and S.P.Victor, “Web Sentiment Analysis for Scoring Positive or Negative words using Tweeter Data”. International Journal of Computer Applications (0975 – 8887) volume 96– No.6, June 2014.

[38] Qingxi Peng and Ming Zhong, “Detecting Spam Review through Sentiment Analysis”. JOURNAL OF SOFTWARE, VOL. 9, NO. 8, AUGUST 2014.

[39] Bing, X., Liang, Z., 2014. “Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training”. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers). Association for Computational Linguistics, Baltimore, Maryland, USA.

[40] Chetan, K., Atul, M., 2014. “A scalable lexicon based technique for sentiment analysis”. Int. J. Foundations Comput. Sci. Technol. 4 (5).

[41] Xia, H., Jiliang, T., Huiji, G., Huan, L., 2014. "Social spammer detection with sentiment information". In: ICDM '14 Proceedings of the 2014 IEEE International Conference on Data Mining.
[42] Duyu, T., Bing, Q., Ting, L., Qiuhui, S., 2014. "Emotion analysis platform on Chinese microblog". CoRR J.
[43] Sujata, R., Parteek, K., 2014. "Challenges of sentiment analysis and existing state of art". Int. J. Innov. Res. Comput. Sci.
[44] Svetlana, K., Xiaodan, Z., Saif, M.M., 2014. "Sentiment analysis of short informal texts". J. Artif. Intell. Res. 50.
[45] Basant Agarwal, Namita Mittal, "Prominent Feature Extraction for Review Analysis: An Empirical Study", In Journal of Experimental and theoretical Artificial Intelligence, Taylor Francis, 2014, DOI:10.1080/0952813X.2014.977830.
[46] Basant Agarwal, Namita Mittal, "Semantic Feature Clustering for Sentiment Analysis of English Reviews", In IETE Journal of Research, Taylor Francis, Vol: 60 (6), pages 414-422, 2014.
[47] Doaa, M.E., Hoda, M.O.M., Osama, I., 2015. "Online paper review analysis". Int. J. Adv. Comput. Sci. Appl. 6 (9).
[48] Basant, A., Namita, M., Pooja, B., Sonal Garg 2, 2015. "Sentiment Analysis Using Common-Sense and Context Information". Hindawi Publishing Corporation Computational Intelligence and Neuroscience.
[49] Yanfang, C., Pu, Z., Anping, X., 2015. "Sentiment analysis based on expanded aspect and polarity-ambiguous word lexicon". Int. J. Adv. Comput. Sci. Appl. 6 (2).
[50] Rajni Singh and Rajdeep Kaur, "Sentiment Analysis on Social Media and Online Review". International Journal of Computer Applications (0975 – 8887) volume 121 – No.20, July 2015.
[51] Fang and Zhan, "Sentiment analysis using product review data". Journal of Big Data 2015.
[52] Lucie Flekova, Eugen Ruppert and Daniel Preot¸ iuc-Pietro, "Analyzing domain suitability of a sentiment lexicon by identifying distributionally bipolar words". Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015), Lisboa, Portugal, 17 September, 2015.
[53] V. Pream Sudha and R. Kowsalya. "A survey on deep learning techniques, Applications and Challenges". International Journal of Advance Research In Science And Engineering, Vol. No.4, Issue 03, March 2015.
[54] Basant Agarwal, Namita Mittal, Pooja Bansal, Sonal Garg, "Sentiment Analysis Using Common-Sense and Context Information", In Computational Intelligence and Neuroscience, Article ID 715730, 9 pages, 2015, (SCIE Indexed), DOI: http://dx.doi.org/10.1155/2015/715730 \.
[55] Doaa Mohey El-Din Mohamed Hussein, 2016. "A survey on sentiment analysis challenges". Journal of King Saud University – Engineering Sciences.
[56] Neeru Mago, 2016. "Opinion mining: Application, techniques, tools, challenges and future trends of sentiment analysis". International Journal of Computer Engineering and Applications, Volume X, Issue IV.
[57] Mr. B. Narendra, Mr. K. Uday Sai, Mr. G. Rajesh, Mr. K. Hemanth, Mr. M. V. Chaitanya Teja, Mr. K. Deva Kumar. "Sentiment Analysis on Movie Reviews: A Comparative Study of Machine Learning Algorithms and Open Source Technologies". Published Online August 2016 in MECS.
[58] Mahmoud Elgamal, 2016. "Sentiment Analysis Methodology of Twitter Data with an application on Hajj season". International Journal of Engineering Research & Science (IJOER).
[59] Anjali Mahavar, Priya Pati, Abhishek Tripathi. "A Survey Paper on Twitter Sentiment Analysis of Current Affairs". IJSRD - International Journal for Scientific Research & Development| Vol. 4, Issue 10, 2016 | ISSN (online): 2321-0613.
[60] Pierre FICAMOS, Yan LIU. "A Topic based Approach for Sentiment Analysis on Twitter Data". International Journal of Advanced Computer Science and Applications, Vol. 7, No. 12, 2016.
[61] Vidisha M. Pradhan, Jay Vala, Prem Balani. "A Survey on Sentiment Analysis Algorithms for Opinion Mining". International Journal of Computer Applications (0975 – 8887) Volume 133 – No.9, January 2016.
[62] Salma Farooq, Hilal Ahmad Khanday. "Opinion Spam Detection: A Review". International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 12, Issue 4, (April 2016).
[63] Lina Maria Rojas-Barahona. "Deep learning for sentiment analysis". Lang Linguist Compass 2016; 10: 701–719.
[64] Yuhai Yu, Hongfei Lin, Jiana Meng and Zhehuan Zhao. "Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks". Algorithms 2016, 9(2), 41; doi:10.3390/a9020041.
[65] Basant Agarwal, Namita Mittal, Springer Book Series: Socio-Affective Computing series, book titled, "Prominent Feature Extraction for Sentiment Analysis", Published by Springer International Publishing, 2016, ISBN: 978-3-319-25343-5, DOI: 10.1007/978-3-319-25343-5, pages: 1-115.
[66] Waheed H, Anjum M, Rehman M, Khawaja A (2017). " Investigation of user behavior on social networking sites". PLoS ONE 12(2): e0169693. doi:10.1371/journal.pone.0169693.
[67] Qurat Tul Ain, Mubashir Ali, Amna Riaz Amna Noureen, Muhammad Kamran, Babar Hayat and A. Rehman. "Sentiment Analysis Using Deep Learning Techniques: A Review". International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.
[68] Gupta, R.K. and Bhargava, S., 2013. A Novel Approach for Cluster Outlier Detection in High Dimensional Data [J]. *International Journal*, *2*(7), pp.131-134.Bhargava, S., Hemrajani, N., Goyal, D. and Gander, S., 2011. DWH-Performance Tuning for Better Reporting. *International Journal of Computer Applications*, *32*(1).
[69] Bhargava,S., Goyal,S. 2013. Dynamic Load Balancing in Cloud Using Live Migration of Virtual Machine. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol. 2 Issue. 8
[70] A. Julka, S. Bhargava"A Static Hand Gesture Recognition Based on Local Contour Sequence", International Journal of Advanced Research in Computer Science and Software Engineerin, vol. 3, no 7, 2013, pp 918-924.