iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** II **Month of publication:** February 2022

DOI: https://doi.org/10.22214/ijraset.2022.40518

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

Sentimental Analysis-A Review

Anil Arora[1], Mr. Subhash[2], Mr. Karmbir[3]
[1]*Department of Computer Science, Gateway Institute of Engineering & Technology (GIET), Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat*
[2]*Department of Mathematics, Delhi Vidya Peeth (DVP), Sonepat*

Abstract: Sentiment can be defined as a view or opinion that is held or expressed. Whereas, sentiment analysis (SA) is a process of computationally identifying and categorizing opinions expressed in a piece of textual content, particularly to decide the writer's attitude towards a particular topic, product or issue. SA also known as opinion mining is widely used in many domains such as products, services, issues and politics to analyze user's behaviors or opinion regarding the related topics. Sentiment analysis is used to analyse the review of product, political issues, reviews on the social media such as face book, twitter, movies etc. where the different users can provide feedback and analyze it. It is also important for Business development by providing the product review and knows exactly what the customer wants. In this work, we have reviewed the latest developments in sentiment analysis.

Keywords: Machine Learning, Supervised Learning, Support Vector Machine, Naive Bayes.

I. INTRODUCTION

Sentimental analysis [1] is the technique of analysing the sentiments of the people or the customers which give their opinion regarding social topics, election polls, about the product on various platforms like social media (facebook, twitter, instagram etc.) or blog writing or feedback form or complain calls to customer care. Through these sources we get numerous amount of information about the flaws in a system or a product which help companies to have an edge over their competitors or to know that where decision taken by authority is liked by the people or not. By selecting appropriate feature like part of speech tagging, senti-word net dictionary and appropriate algorithms for classification, we can get improved results and hence improved accuracy is achieved by combining the techniques. It is a very challenging task because there is a lot of data available for its analysis and that too in many forms so there is a need of efficient classification method that will not only classify the data but also provide accurate results in optimum time.

Sentiment Analysis is sentiment classification that uses Natural language processing, Text analysis or contextual mining for the extraction and classification of the sentiment on the basis of sentiment reviews. There are various classification techniques included in machine learning such as support vector machine, naïve bayes, k-nearest neighbour etc. that is used to classify the sentence or a statement i.e. indicate positive , negative and neutral or that can be represented as on n-scale point. Sentiment Analysis is also referred as opinion mining.

Sentiment Analysis could be used in different fields like politics, Business, personal , Social media, Data Analytics, Data Science, Information retrieval (Cognitive Weight).It can be categorized in various forms like good, very good, bad, better, statisfcatory, very bad which depicts a emotion of user in business or a citizen of a country involve in voting of Election. In economic matters it helps the companies to improve the quality of the product and to know the estimated extend of their product and its acceptance .In Social Media, the sentiment analysis can be seen in the form of posts, comments, followers or put on by people in facebook, twitter, instagram etc. here they express their opinion on a specific subject which is connected to their sentiment too [2].

II. CLASSIFICATION OF SENTIMENTAL ANALYSIS

Sentiment Classification techniques can be roughly divided into machine learning approach, lexicon based approach. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The text classification methods using ML approach can be roughly divided into supervised and unsupervised learning methods. The supervised methods make use of a large number of labeled training documents. There are many kind of supervised classifiers in literature. [3][4].

There are various supervised classification methods among which any can be used within sentiment analysis method as it the part of text classification problem. A classifier that comprises of Bayes theorem is known as Naïve-Bayes classifier. It is a simple probabilistic classifier in which it is assumed that the presence or absence of one feature in a document is not reliable on any other feature present within that document.

©IJRASET: All Rights are Reserved

A. Machine Learning

Machine Learning is the core of various future restricted advancement in the world .Today we can see examples of Machine Learning such as Sophia AI robot, Self driving car, Tesla etc. Machine learning is also a subset of artificial intelligence. It focuses mainly on the design of systems, thereby allowing them to learn and make predictions based on the basis some experience .Machine Learning is based of idea that we give access to machine and learn from it and also extraction pattern from data i.e. used to find optimal behavior. Example is Optimize Speech Recognition used in Apple I phone Siri is how the functioning of the siri works a powerful speech recognition kicks off and convert into audio corresponding text form which is then send to the server for further processor then neural language processing algorithm are run to understand user experience and final answer is told by the siri.

B. Supervised Learning

Supervised Learning works under supervision in which learning it train the data which is well labeled and predict with the help of the label data set. A labeled data set is a set which we already the output to its corresponding input .Supervised Learning is defined as the X as the input and Y as its output. We have image as spoon and knife and feeded to machine analysis and learn its association with its label feature such as sharpness , size etc. and new image is send to machine without any label and recognize either knife or spoon according its features. Algorithm teaches moral to learn from label .Supervised learning can be classified into classification and regression.

C. Support Vector Machine

SVM is a supervised machine learning algorithm capable of dealing with the highly complex data .It is used to model on n-dimensional problems which divides the training data into two classes efficiently using a frontier called hyper plane or line. It is used for both classification and regression problems. It is used in classification problems. In this algorithm , we plot each data item as a point in n-dimensional space(where n is number of features) with the value of each feature corresponding to n coordinate .There we perform classification by finding the hyper-plane that differentiates the two classes efficiently

D. Probabilistic Classifiers

Probabilistic classifiers use mixture models for classification. The mixture model assumes that each class is a component of the mixture. Each mixture component is a generative model that provides the probability of sampling a particular term for that component. These kinds of classifiers are also called generative classifiers. Three of the most famous probabilistic classifiers are discussed in the next subsections.

E. Naive bayes

The Naive Bayes classifier is the simplest and most commonly used classifier. Naive Bayes classification model computes the posterior probability of a class, based on the distribution of the words in the document. The model works with the BOWs feature extraction which ignores the position of the word in the document. It uses Bayes Theorem to predict the probability that a given data set belong to a particular label.

$$P(label|features) = \frac{P(label) * P(features|label)}{P(features)}$$

P(label) is the prior probability of a label or the likelihood that a random feature set the label. P(features|label) is the prior probability that a given feature set is being classified as a label. P(features) is the prior probability that a given feature set is occurred. Given the Naive assumption which states that all features are independent, the equation could be rewritten as follows:

$$P(label|features) = \frac{P(label) * P(f1|label) * \ldots .. * P(fn|label)}{P(features)}$$

An improved Naïve Bayes classifier was proposed by Kang and Yoo to solve the problem of the tendency for the positive classification accuracy to appear up to approximately 10% higher than the negative classification accuracy. This creates a problem of decreasing the average accuracy when the accuracies of the two classes are expressed as an average value.

F. Bayesian Network (BN)

The main assumption of the NB classifier is the independence of the features. The other extreme assumption is to assume that all the features are fully dependent. This leads to the Bayesian Network model which is a directed acyclic graph whose nodes represent random variables, and edges represent conditional dependencies. BN is considered a complete model for the variables and their relationships.

Therefore, a complete joint probability distribution (JPD) over all the variables, is specified for a model. In Text mining, the computation complexity of BN is very expensive; that is why, it is not frequently used.

G. Maximum Entropy Classifier (ME)

The Maxent Classifier (known as a conditional exponential classifier) converts labeled feature sets to vectors using encoding. This encoded vector is then used to calculate weights for each feature that can then be combined to determine the most likely label for a feature set[11]. This classifier is parameterized by a set of X{weights}, which is used to combine the joint features that are generated from a feature-set by an X{encoding}. In particular, the encoding maps each C{(featureset, label)} pair to a vector. The probability of each label is then computed using the following equation:

$$P(fs|label) = \frac{dotprod(weights, encode(fs, label))}{sum(dotprod(weights, encode(fs, l)) for l in labels)}$$

ME classifier was used by Kaufmann to detect parallel sentences between any language pairs with small amounts of training data. The other tools that were developed to automatically extract parallel data from non-parallel corpora use language specific techniques or require large amounts of training data. Their results showed that ME classifiers can produce useful results for almost any language pair. This can allow the creation of parallel corpora for many new languages.

III. LITERATURE REVIEW

Various types of sentimental analysis techniques are available. In this section, we provide the literature review of work done in this field.

Jia et al., [5] proposed a new method of semantic similarity calculation. The concepts were classified into three classes: simple concept; complex concept and combined concept. To different concept, different method was designed and then transformed the similarity calculation of concept into the similarity calculation of the sememe. The similarity of the sememe was computed by the hyponymy of the sememe in the sememe tree. Experiments showed the new approach was effective to the similarity calculation and outperformed the conventional computed approaches.

Schneider et al. [6] proposed a novel matrix learning strategy for extending relevance learning vector quantization (RLVQ), an effective prototype-based classification protocol, toward a general adaptive measure. Through introduction of a full matrix of relevance factors in the distance metric, correlations between various attributes as well as their significance for classification occurs at the time of training. When contrasted with weighted Euclidean measure utilized in RLVQ as well as its variants, a complete matrix is more powerful for representing the internal structure of data adequately. Huge margin generalization bounds may be transferred to the case resulting in bounds that are not dependent on input dimensionality. This is true for local measures attached to all prototypes that correspond to piece-wise quadratic decision bounds. The protocol was evaluated in contrast to alternate LVQ strategies through usage of artificial dataset, a benchmark multiclass issue from UCI repository, as well as a problem from bio-informatics, the recognition of splice sites for C.

Jiang et al. [7] proposed a sentiment mining and retrieval system which mines useful knowledge from product reviews. Furthermore, the sentiment orientation and comparison between positive and negative evaluation were presented visually in the system. Outcomes of experiments on a real-world dataset have shown the system is both feasible and effective.

Nielson et al. [8] discussed the particular challenges of sentiment analysis in the domain of social media messages. Its purpose is a rule-based method with constructing a shallow linguistic analysis containing named entity extraction and event recognition. It works for producing a sentiment polarity and score for a given tweet.

Kumar et al. [9] proposed rule-based method for entity-level sentiment analysis in Twitter. They evaluated a sentiment score for each entity depending on its textual proximity to words from a sentiment lexicon. It also executed simple anaphora resolution by resolving pronouns to the closest entity in the tweet. The rule-based algorithm differentiates between demonstrative, imperative and interrogative sentences and can, among other things, handle comparative sentences, negation and but-clauses. For enhancing the recall of the proposed approaches, the researchers recognize extra tweets that are likely to be opinionated and train a support vector machine (SVM) to appropriate polarity labels to the contained entities.

Omar et al. [10] focused on reducing the number of features in dataset by selecting only the relevant features before giving the dataset to classifier. This motivated the need for sufficient methods that capable of selecting the relevant features with minimal information loss. The aim was to reduce the workload of classifier by using feature selection methods. With the focus on classification performance accuracy, the concept was highlighted, abilities and application of feature selection for various applications in classification problem. From the review, classification with feature selection methods has shown impressive results with significant accuracy when compared to classification without feature selection.

Veeraselvi et al. [11] presented opinion detection and organization subsystem, which had already been integrated into proposed larger question-answering system.

The subjectivity classification system used Genetic-Based Machine Learning (GBML) technique that considered subjectivity as a semantic problem. The classification of a review was estimated through the average semantic orientation of phrases in the review which comprise adjectives or adverbs. Experimental results of the proposed techniques were efficient and generated eminent evaluations.

Doaa Moha et. al. [12] in their paper included the sentiment analysis that evaluates challenges in sentiment based on comparison between sentiment analysis review structure and sentiment analysis challenges. This comparison out another major faster in concept of sentiment analysis. A negative phase of this challenge is that something it may differ in implicit and explicit meaning of a review. This could lead to misinterpretation of particular review.

The second comparison based sentiment analysis challenges relevant to accuracy rate resulting in evaluation of sentiment and selecting sources to improve accuracy. The theoretical type of sentiment technique could be used for solving sentiment challenges emphasis on average of accuracy based on number of research in each challenge. More the research in sentiment challenge less the average accuracy rate.

Soujana Poria et al. [13] described aspect the extraction that identifies, the main aim is opinionated text which could be either in form appreciation or compliment and complaint. The tag each word opinionated sentence as aspect or non-aspect , a seven layer deep convolution network along with set of language patterns is used which results in classifier .Obtain better accuracy and complied with the word embedding model for sentiment analysis.

Ankit Kumar Soni [14] described that There is no such technique proposed which can help in handling the multil anguage data. In this paper, Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm. Amongst various algorithms, the results are compared which can help in analyzing the performance of various algorithms amongst each other and show which has provide to be better. It is seen through the results achieved that the proposed technique has provided better results in comparison to other existing approaches.

Salloum et. al [15] declared that the polarity of a text can be determined by two approaches either by machine learning or by lexicon approach. The classifiers used were Naïve Bayes (NB), Support Vector Machine(SVM) and K-Nearest Neighbor(KNN) in which highest precession was given by Support Vector Machine and highest recall by K-Nearest Neighbor. For testing of data set cross validation (10 -fold) was used with the best precision of 75.25 by Support Vector Machine and best recall of 69.04 by Naïve Bayes. For better classification it demands use of bigger data sets which can be labeled by crowd sourcing followed by semi supervised learning.

Neha Rajput et. al. [16] described that Any kind of attitude, through or judgment that occurs due to any feeling is known as a sentiment which is also known as opinion mining. The sentiments of individuals towards particular elements are analyzed in this approach. To gather sentiment information, web or internet is the best known source. A platform that is accessed socially by various users to post their views is known as Twitter. The messages that are posted by these users are known as tweets. The properties of Tweets are highly unique due to which new challenges have raised. In comparison to several other domains, the sentiment analysis requires higher analysis studies.The sentiment analysis is the technique which is applied to analyze sentiment. The sentiment analysis techniques has various phases which are data collection, data cleaning, and classification. In this paper, various sentiment analysis techniques are review and analyzed in terms of certain parameters.

IV. CONCLUSION

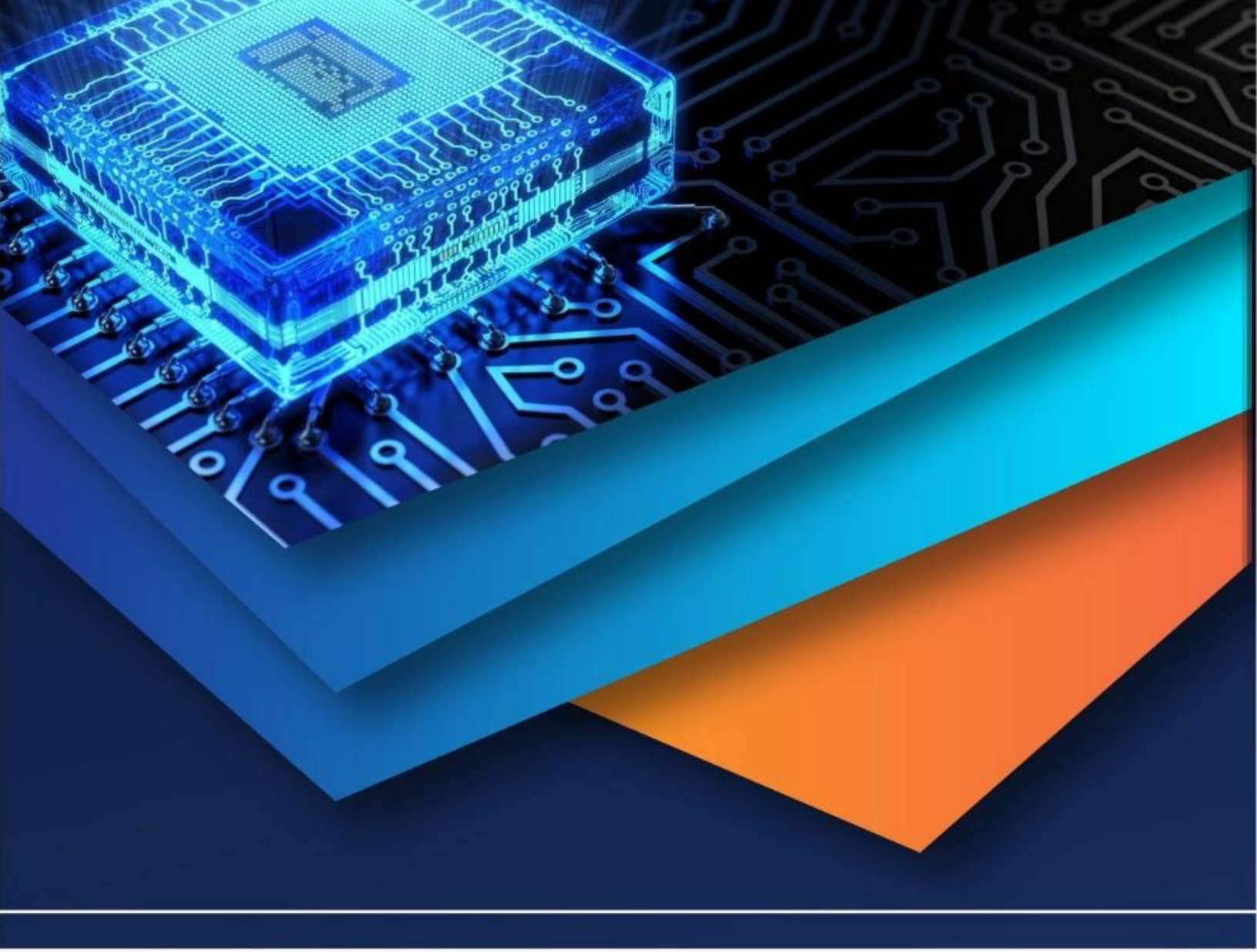
Sentiment analysis is a field which is catching up in the recent years and its applications are subject to increase to a broader range in near future. This work is an attempt to create a basis with the help of which future works can be improved and also take a note of the challenges this field offers. In this work, the latest developments in sentiment analysis are reviewed and the future possibilities for each of these developments are presented. The effectiveness of various approaches has been evaluated and shown.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Chetashri Bhadane,Hardi Dalal, Heenal Doshi, 2015, "Sentiment Analysis: Measuring Opinions", Science Direct, Volume-45.
[2] Cambria E., Schuller B., Xia Y. and Havasi C, 2013, " New avenues in opinion mining and sentiment analysis".
[3] Walaa Medhat a,*, Ahmed Hassan b, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal (2014)
[4] Gautami Tripathi and Naganna S, 2, June 2015, "Feature Selection And Classification Approach For Sentiment Analysis", Machine Learning and Applications: An International Journal (MLAIJ) Volume- 2.
[5] Keliang Jia, Jibin Fu, "Semantic similarity computation based on HowNet2008", Natural Language Processing and Knowledge Engineering, 2008.
[6] Petra Schneider, Michael Biehl, Adaptive Relevance Matrices in Learning Vector Quantization, *in* Neural Computation 21:3532-3561 · September 2009
[7] Jiang P., Zhang C., Fu H., Niu Z. and Yang Q., December, "An approach based on tree kernels for opinion mining of online product reviews. In Data Mining" , IEEE 10th International Conference on IEEE, 2010.
[8] Nielsen, F. A. "ANEW: Evaluation of a word list for sentiment analysis in microblogs". In Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages, volume 718 of CEUR Workshop Proceedings, pages 93–98, 2011.
[9] Kumar, A., & Sebastian, T. M., "Sentiment Analysis On Twitter". IJCSI International Journal of Computer Science, Issues 9(3):372–378, 2012.
[10] Omar, N., Jusoh, F., Ibrahim, R., & Othman, M. S. (2013). Review of Feature Selection for Solving Classification Problems. JISRI
[11] S. J. Veeraselvi and C. Saranya, March 2014, "Semantic orientation approach for sentiment classification", International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE).
[12] Doaa Moha, El-Din Mohamed Hussein, "A survey on sentiment analysis challenges", Journal of King Saud University-Engineering Sciences (2016)30,330-338
[13] Soujana Poria, Erik Cambria, "Aspect extraction for opinion mining with a deep convolutional neural network", International Joint Conference on Neural Networks (IJCNN), 2016.
[14] Ankit Kumar Soni, "Naïve Bayes and Maximum Entropy classifiers", International Journal of Advance Engineering and Research Development,2017.
[15] Salloum, S. A., Al Hamad, A. Q., Al Emran, M., & Shaalan, K., "A Survey of Arabic Text Mining", In Intelligent Natural Language Processing: Trends and Applications (pp.417-431). Springer, Cham. 2018
[16] Neha Rajput, Mrs.Shivani Chauhan, "Analysis of various sentiment analysis techniques", International Journal of Computer Science and Mobile Computing (2019).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)