*International Journal on Emerging Technologies* **10**(2): 15-18(2019)

**ISSN No. (Print) : 0975-8364**
**ISSN No. (Online) : 2249-3255**

# Investigation of Various Techniques and Classification Methods on Cognitive Sentimental Learning

***D. Venkatesan, K. Vithiya Ruba and K.R. Sekar***
*School of Computing,*
*SASTRA Deemed University (Tamil Nadu), India.*

*(Corresponding author: D. Venkatesan,)*

*(Received 03 April 2019, Revised 01 June 2019 Accepted 03 July 2019)*
*(Published by Research Trend, Website: www.researchtrend.net)*

**ABSTRACT: Sentiment analysis is done in every field and it is adopted for analyzing the people's mood and also about people's view on products, online sites, etc. The paper deals with how sentiment analysis is important and also gives the detailed analysis of different types of methods involved in sentiment analysis and gives what are all the things to be taken into sentimental analysis had become a part of every-day life. Many techniques have been introduced for analyzing the sentiment and the accuracy of the result fully depends on the type of the application that are working with and also on what Application programming Interface (API) and analyzing the sentiments for different applications. Further improvements in techniques may occur in future for sentimental analysis.**

**Keywords**: sentiment analysis, sentiment, opinion, sentiment tool, polarity

## I. INTRODUCTION

Sentiment analysis [1] has many different sources like Online site reviews, Micro-blogging. Many different methods are followed to determine the polarity but the common methods in use are lexicon based and machine learning based techniques. This paper also deals with how to create our own sentiment analysis tool which helps in detecting the polarity.

Machine learning can be classified into two types: supervised and unsupervised learning, whereas these two types of learning based techniques has many algorithms which can be used for performing the analysis [3-4].

Sentiment analysis tool for Twitter and Facebook [5] can be built and also many programming languages supports in developing the tool. Usually sentiment analysis done through programming languages involves the steps like collecting the data, labeling the data, preparing the data and training our prepared data and at last making predictions on our analysis [6-7].

But many online sentiment analysis tools are also available, but the accuracy will not be up to our expectations and also the result produced will not give the correct analysis of the products or reviews. So, sentiment analysis with our own effort gives the better result.

## II. RELATED WORKS

An approach for detecting the sentiment analysis on noisy and biased data [2], is given where the massages are classified as subjective and objective further the subjective messages are distinguished as positive and negative, the classifier used here is a supervised classifier which reduces the effort on labeling. Noisy and biased labels can be used as training data, though they have different bias an effective result was achieved by combining them.

A method of using the data sets for micro-blogging like hash tag, emotion and I Sieve and features like n-grams, lexicon, part-of-speech [7] is used. Here, the combination of the features n-grams, lexicon and micro-blogging on hash tagged dataset outperforms other features; emotion dataset gives an improvement in the absence of a micro - blogging feature.

A sentiment analysis based on subjectivity [8] is used. They proposed a method based on machine learning to detect the subjective portion of the sentence using text-based classification technique and hence minimum cuts in the graph can be identified.

A method to perform sentiment analysis for an informal text like SMS and the negated words are taken into account has been proposed. For negated words separate dictionary called sentiment lexicon dictionary is created and hence the performance gain is up to 6.5 when generated lexicons are used [10].

A custom Sentiment analysis tool for Twitter post [12-13] has been discussed which increases the performance by increasing the overall scoring of tweets compared to the third party result.

## III. IMPORTANCE OF SENTIMENT ANALYSIS

### *A. Techniques for Sentiment Analysis*

The following techniques are to be considered for sentiment analysis without these techniques successful sentiment analysis cannot be done.

**Lexicon Based vs Machine Learning.** When Lexicon Based technique [8-9] is followed, it is based on the dictionaries where the words are listed in the dictionary and is annotated with both the polarity and the strength for calculating the polarity of the sentence and also this gives high precision and low recall. But when Machine learning[4] is used we have to label the given example and then train the algorithm with the given example. Learning based technique is better when compared to Lexicon because in Lexicon based technique we have to depend on the dictionary which will not be available for all the languages and also in case of learning based technique it requires both labeling and training.

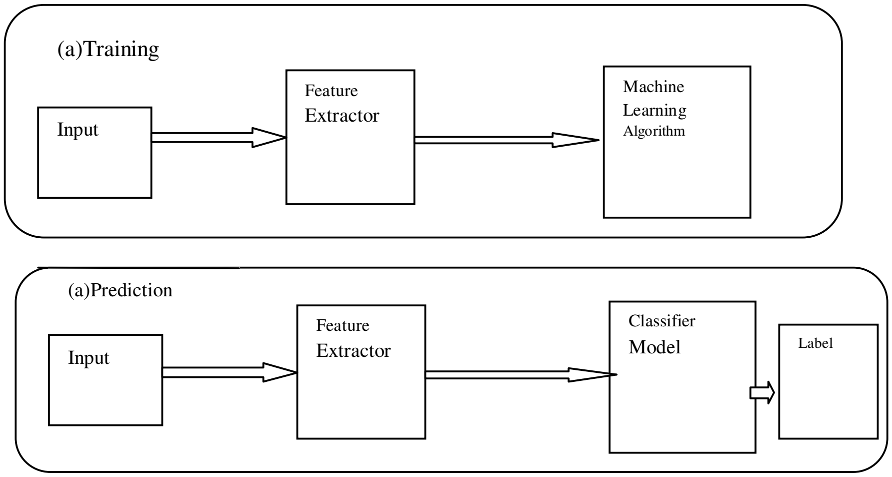

**Fig. 1.** Example for Machine Learning.

According to the Lexicon Based formula the accuracy was calculated.

$$\text{pol}(T) = \sum_{i=1}^{k} \text{pol}(m_i) \qquad T=\{m_1....m_k\} \qquad (1)$$

$$\text{pol}(m_i) = \sum_{j=1}^{n} \text{score}(t_j) \qquad M=\{t_1...t_n\} \qquad (2)$$

**Statistical vs Syntactic Techniques.** By general, syntactic technique gives the accurate result when compared with the statistical technique because syntactic technique follows syntactic rules and also it helps in detecting the verbs, adjectives and nouns, but one of the drawbacks is that it depends on the languages for which we are performing the sentiment analysis and also the classifiers result cannot be ported to other languages. On the other hand, statistical technique [11] is fully based on the relationship between words and categories, but we can use it on any other language and when the machine translation of the original data set is used it gives good results

**Neutral Class.** Neutral class is very important for detecting the accurate sentiment, because when an SVM classifier is used, its overall accuracy is said to be increased because the neutral words are also classified as positive and negative moreover it will lead to over fitting. So, neutral class is also equally important like positive and negative class.

**Tokenization Algorithm.** In the tokenization algorithm, n-gram framework is said to be used and the value of n should not be too big. 2-grams or 3-grams can be comfortably used and also the combinations of the keyword should be reduced and also the occurrences of the word should not be minded whereas when Binarized versions is used for the multiple occurrences of the words then better accuracy is achieved.

**Feature Selection Algorithm.** Usually the words or features in the word is extracted which can be used in the tokenization algorithm and we cannot use all the words because there are many irrelevant words which are not needed for our analysis. Mutual Information and Chi-square test are the two mainly used feature selection methods. The result and also the accuracy of the algorithm depends on the type of the application we are analyzing and the different configurations used, whereas we have to go for trial and error method to find which method and configuration suits the algorithm to give the best result.

$$\text{IG}(t) = -\sum_{i=1}^{m} p(c_i)\log p(c_i) + p(t)\sum_{i=1}^{m} p(c_i \mid t)\log p(c_i \mid t) + p(\bar{t})\sum_{i=1}^{m} p(c_i \mid \bar{t})\log p(c_i \mid \bar{t}) ------- (3)$$

**Classifiers may deliver different results.** We have to make sure that all the classification methods are tried because when feature selection is used some of the classifiers may outperform. Usually SVM is said to outperform all other classifiers, but Naïve Bayes[3] is also said to perform well or even gives the better result than other outperforming algorithms so we should not eliminate the classifiers [14-15].

**A) Types of Classifiers**

**Bayes** – Always the optimal (minimum error rate or minimum risk) but requires exact knowledge of class prior probabilities and class conditional probabilities of features. Seldom possible because exact nowledge rarely exists.

**Bayes linear** – Assumes Gaussian distribution of features with equal covariance matrices for each class. A modest number of parameters to estimate. Fast training and classifying. In general, performance is limited.

**Bayes quadratic** – Assumes Gaussian distribution of features with a separate covariance matrix for each class. Requires many parameters (feature covariances) to be estimated. Fast training and classifying. Performance may be poor when data is significantly non-Gaussian.

**Nearest neighbour** (1-nearest neighbour) – A simple nonparametric method that uses all the training data for classification. Has high computational complexity for classification, though some acceleration methods exist. Must select a metric. Upper bound on error rate approaches twice that of ideal Bayes classifier.

**k-Nearest Neighbour** – A robust non-parametric classifier. Classification has high computational complexity when. Must select metric and value of k. k must be set using validation. Can have excellent performance for arbitrary class conditional pdfs.

**Parzen window** – Robust non-parametric. Must select form of kernel and size parameter h. Complexity and performance is similar to k-NN method.

**Neural network** – The multi-layer perceptron (a non-parametric classifier) is the standard network to use for supervised learning. Other types of neural networks are useful for unsupervised learning. Training can be very slow, but classification is fast. The number of hidden nodes must be set using validation. Can have excellent performance. Impossible for a human to "understand" the classifier. Performance is vulnerable to unforeseen input data.

**Decision tree** – non-metric method. Gives a set of rules that can be understood.

**The domain or topic is important.** What domain we are going to discuss also matters a lot because the accuracy of the classifier depends on the domain some classifier may give 90% accuracy in one domain which may vary it to 60% in some other domain so it depends on the type of the classifier we are using when we considered for example the twitter application lexicon based techniques should be avoided because users may use idioms, jargons which may heavily affect the polarity.

**Techniques may vary.** Some techniques may vary and it may fail to work on some domains and also the quality of the result may not be good so, we should not depend on the particular technique which may result in a poor result and also the algorithm may be unnecessarily complicated.

One of the important techniques for building a highly accurate classifier is to use the ensemble learning where we can combine the ensemble learning results and the results of the different classifiers which can be presented in the different dimensions like 2D and 3D but unfortunately the ensemble learning may fail in the text analysis and so the ensemble learning is avoided in most cases.

## IV. API's IN SENTIMENT ANALYSIS

This section discusses about the importance of sentiment analysis and it is such an essential thing for small businesses which provides the offer to improve their business and helps in discovering their unique quality about their products. It will help them to identify their customer needs.

There are many API's available for sentiment analysis like Semantria API, Alchemy API, Text processing API, NLP Tools API, Sentiment analysis API, etc., these API can be available for both cost and free.

Among which Alchemy API uses machine learning and also it uses a Natural Language Processing which involves both the sentiment analysis and semantic text analysis which helps the clients over the cloud. This Alchemy API is useful in analysis of document- level, entity- level sentiment, keyword- level sentiment for monitoring the social media to analyze the trend. This Alchemy API returns the result in the form of XML, JSON and also in RDF formats. Many different extractions are used like relation, text, face, image link, author, Language Detection. Here the API calls include HTML API, Text API, Web API. It supports the languages like English, French, German, Italian, Portuguese, Spanish, Russian.

Some kind of API's can be used in the programming languages, but there are some API's which can be used as tools. One of the API tools is Sentiment 140. Sentiment 40 is a recommended tool for classifying the polarity of tweets. Just by registering for sentiment 140 API, we can easily make use of the API as a tool, but the accuracy will not be up to our expectation.

Semantria API is used for text or sentiment analysis in the cloud which is very fast, distributed, scalable and highly customizable. Semantria cloud is said to be well maintained for all the users, which helps the users in running on both their desktop and mobile. This API also supports languages like C++, JAVA, PHP, .NET, Python, Ruby and Javascript with the help of these languages the implementation can be made to the best.

There are several online API's which can be used without any cost one of the examples is Viral Heat API which helps in analyzing the social media just by signing up for free it was already used in beta and also three analytics partners are included like Zuberance, Seesmic and Klout.

Thus, many API's are available which can be used by the developers and it depends on the users to use free API or Cost API

## V. RESULTS AND DISCUSSIONS

In the sentimental analysis accuracy for the methodologies were appreciated more for the past one decade. In the work, the proposed technique gives high precision of results in terms with Quality of Service (QoS) accuracy. The dataset records taken in the Table 1.were tested with WEKA tool to know the accuracy for the same dataset applied on different methods in sentimental analysis. Reports are exhibited in the paper below.

**Table 1. Sentimental analysis Report.**

| Methodologies | DataSet Records | Accuracy Rate |
|---|---|---|
| Bayes Theorem | 944 | 82.4 |
| KNN | 944 | 72.9 |
| SVM | 944 | 80.7 |
| Syntactic Techniques | 944 | 79.7 |
| Ensemble model | 944 | 85.4 |

## VI. CONCLUSION

In the sentimental analysis various techniques were applied so for by the researchers. In the paper, plenty number of methodologies were proposed like Machine learning algorithm, Syntactic Techniques, Support Vector Machine and Feature selection for sentimental analysis for different applications that gives the long mileage and mile stone to the customers to get good products in online shopping's. The other techniques like Bayes, Neural Networks, K-NN and Natural Language Processing also used to support the above methods to attain the perfect result in sentimental analysis.

## REFERENCES

[1]. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R. (2011). Sentiment analysis of twitter data, *In Proceedings of the ACL 2011 workshop on languages in social Media*, 30-38.

[2]. Barbosa, L., Feng, J. (2011). Robust sentiment detection on twitter from biased and noisy data, *In Proceedings of the 23rd international conference on computational linguistics,* 2010, 36-44.

[3]. Claster, W.B., Cooper, M., Sallis, P. (2010). Modeling sentiment from twitter tweets using Naïve Bayes and unsupervised artificial neural nets, *In Proceedings of the CIMSIM'10,* 2010, Sept., 28-30.

[4]. Gharehchopogh, F.S., Khaze, S.R., Maleki, I. (2015). A New Approach in Bloggers Classification with Hybrid of K-Nearest Neighbor and Artificial Neural Network Algorithms. *Indian Journal of Science and Technology,* 2015, Feb., **8**(3), 237-246.

[5]. He, Y. and Alani, H. (2011). Semantic smoothing for twitter sentiment analysis, *In Proceedings of the10th*

*international semantic web conference (ISWC)*, 2011, Oct., 23-27.

[6]. Iwanaga, I., Nguyen, T.M., Kawamura, T., Nakagawa, H., Tahara, Y., Ohsuga, A. (2011). Building an earthquake evacuation ontology from twitter*, In Proceedings of the IEEE international conference on granular computing (GrC), 2011,* Nov. 8-10, 306–311.

[7]. Kouloumpis, E., Wilson, T. and Moore, J. (2011). Twitter sentiment analysis: The good the bad and the OMG. *In Proceedings of the ICWSM*, 2011, 538-541.

[8]. Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval*, 2008, Jan., **2**(1-2): 1-135.

[9]. Zol, S. and Mulay, P. (2015). Analyzing Sentiments for Generating Opinions (ASGO)-A New Approach. *Indian Journal of Science and Technology,* **8**(S4), 206-211.

[10]. Kiritchenko, S., Zhu, X., Mohammad, S.M. (2014). "Sentiment Analysis of Short Informal Texts", *Journal of Artificial Intelligence Research,* **50**, 723–762.

[11]. Vigneshwari, S., Aramudhan, M. (2015). Social Information Retrieval Based on Semantic Annotation and Hashing up on the Multiple Ontologies. *Indian Journal of Science and Technology*, **8**(2), 103–107.

[12]. Vithyiya Ruba, K. and Venkatesan, D. (2015). Building a Custom Sentiment Analysis Tool based on Ontology for Twitter Posts, **8**(13): 1.

[13]. Tantray, Muneeb Ahmad and Parveen (2018). Traffic Congestion Analysis of High Volume Road Stretches in Srinagar and Pulwama, *International Journal on Emerging Technologies*, **9**(2): 01-06.

[14]. Manikandan, G., Bala Krishnan, R., Preethivi, E. Sekar, K.R., Manikandan, R. and Prassanna, J. (2019). An Approach with Steganography and Scrambling Mechanism for Hiding Image over Images, *International Journal on Emerging Technologies*, **10**(1): 64-67.

[15]. Monica, B., Rajkumar, K., Sekar, K.R., Manikandan R. and Bagyalakshmi, K. (2019). Cognitive Knowledge of Routing Protocol Configuration in Smart City. *International Journal on Emerging Technologies*, **10**(1): 59-63.

**How to cite this article:** Venkatesan, D., Vithiya Ruba, K. and Sekar, K.R. (2019). Investigation of Various Techniques and Classification Methods on Cognitive Sentimental Learning. *International Journal on Emerging Technologies*, **10**(2): 15-18.