Sentiment Analysis: Methodologies and Approaches

[1]Er. Amandeep Kaur, [2]Dr. Amardeep Singh

[1]M.Tech Research Scholar, [2]Professor
Department of Computer Engineering,
Punjabi University, Patiala

Abstract : Sentiment Analysis is revolutionizing many industries like ecommerce, entertainment, politics etc. at rapid pace. This is an application which belongs to the text based mining and it works in collaboration with machine and deep learning algorithms. Sentiment Analysis helps in increasing productivity of the business along with the better customer experience as it brings meaningful information from unstructured data. This paper includes explanation about sentiment analysis, various techniques and approaches used for it along with naïve bayes algorithm. Various levels of sentiment analysis like document, sentence or aspect based is also described in the paper.

Index Terms - **Lexicon, Machine Learning, NLP, Semantic Analysis, Keyword Spotting**

INTRODUCTION - Twenty first century is a technology driven era where industries around the world are rapidly advancing with the large impact of computer science and internet related technologies. Last some years have seen some revolutionary technologies like Cloud Computing, Internet of Things, Software Defined Networks, Artificial Intelligence, Containerization etc. which are changing the industries and businesses. Machine Learning and Deep Learning is also one of the most powerful technologies came out in recent years. Text Mining is one of the application of machine learning and deep learning. With text mining under machine and deep learning algorithms, computing machines are made to learn structured or unstructured data on which some particular functions can be performed. Making machines think like humans is one thing which is very complex, where neural networks can be used along with deep learning algorithms in order to learn, understand and implement actions without any human interference. Lots of industries have started to use these technologies in the fields like automobile industry with Self Driving Cars, Text Mining from Opinions or Reviews, Security etc. Ecommerce Industry giants like Amazon, Alibaba, Ebay, Flipkart etc. uses text based mining to know about the product reviews and customer sentiments about products. Text Semantics is also used in order to know about the text meaning, where different words are integrated to create a sentence. Words have contextual and lexical relations among them and it further takes towards hierarchy of relationships. Semantics is important to analyze the different relationships among data and get the meaningful information. Semantics need not to have any specific structure of data and it related with the context and meaning, but we sometimes also need to differentiate things like "Apple is a technology related company" from "Apple is a fruit".

2. SENTIMENT ANALYSIS

Sentiment Analysis is an text based mining application which is used in collaboration with many web applications, softwares, mobile apps. The work of sentiment analysis is to extract or analyze sentiments from large dataset of structured or unstructured data. There are different types of text resources which can be used Facebook Comments of some page or individual posts, Tweets and comments in twitter, Google Opinions, IMDB Movie reviews[2-3] etc. These can also be Celebrity pages or tweets, some organization which is selling or endorsing its product on social profiles, political party or politician page or tweets etc. This technology has a major impact on ecommerce industry with online websites like Amazon, or Flipkart etc. have reviews section on all the products page, where customers who have bought the product can review the product which can be helpful in buying or selecting the product out of many available[1] on the basis of customer reviews and ratings. These reviews helps the customers who are willing to buy the product to better know about the product as if the reviews are positive, negative or neutral. The major work of sentiment analysis is to analyze the text data and then to get and understand the review and the sentiment shown by looking at the factors like mood or modality. This mainly is done when the content is subjective, its because with the objective context of data, there is no sentiment expressed and it does not include any sort of emotion or feeling. Text data is basically divided into two types i.e. Subjective, which can be review based and objective, which is totally related with facts. Sentiment Analysis can be extracted using NLP, lexical resources, machine and deep learning algorithms and then the data collected can be used to get the polarity of the document. Polarity of the document can be expressed in either positive, negative or neutral manner. Advanced emotions or expressions like sadness, anger etc. can also be fetched. Various levels of sentiment analysis can be document, sentence or entity based. Document level fetches the sentiment of entire document, but the document has to be based on a single topic. Sentence level extracts the emotion at the sentence level and it looks for the sentiment in every sentence. Sentence with no review has a neutral view. Entity or aspect based level is the detailed analysis. In this form, we are looking at the aspect of the text. For example, in a smartphone review by the customer, "Camera quality is good, but memory is low." Aspects are Camera Quality and Memory in the example. In aspect based, sentiment analysis collaborated two different tasks in single entity level. It fetches the aspect and it then classifies the textual data according to that aspect.

3. METHODOLOGIES

In largest number of cases, text analysis is done at the document level. During polarity analysis, scores are assigned to the levels of expressions i.e. scores are assigned to find the levels of positivity, negativity in the emotions or feelings expressed in the document and a document is accessed as a whole on the basis of the summarized score.

Name	Description	Applications
Keyword Spotting	In this way, text is categorized on the basis of some unambiguous words which can be the part of the sentence or the document. These words can have some special meaning in respective to sentiment analysis.	**Speech Processing, Image Processing, Text Analysis, Sentiment Analysis**
Lexical Affinity	It provides a probabilistic similarity to some words for some particular emotion or feeling.	**Text Analysis, Sentiment Analysis**
Statistical Method	This method calculates the target of affective keywords. It also calculates the co-occurrence numbers on the basis of large training corpus.	**Text Analysis, Sentiment Analysis**

Sentiment Analysis is categorized into three major categories i.e. Lexicon based, Machine Learning Based and the integration both, which is also known as Hybrid Approach. Machine Learning using the mixture of ML algorithms and linguistic features. Lexicon method is dependent on sentiment lexicon which can be a dictionary, or a book containing words. Lexicon in my work can be a dictionary of words which is used to analyzing sentiments as all the words in the dictionary are associated with some special scores which integrated with the positioning of the words in a sentence or a document are used to analyze the sentiments in the document. Sanjida Akter and Muhammad Tareq Aziz[3] have shown how to predict the sentiment of the status post which is regarded as a unstructured dataset. Lexicon Based Approach works best in this kind of works

behind a status post of Facebook which in nature of unstructured dataset, cross language domain and noisy. Traditional opinion mining is not efficient to gather sentiments from social media giant like Facebook and lexicon based dictionary approach works efficiently in such kind of works.

Hybrid Approach is the composite type integrating both machine learning approach and lexicon based approach. Sentiment Analysis further division and sub categories of categories are defined in Figure 1.1:
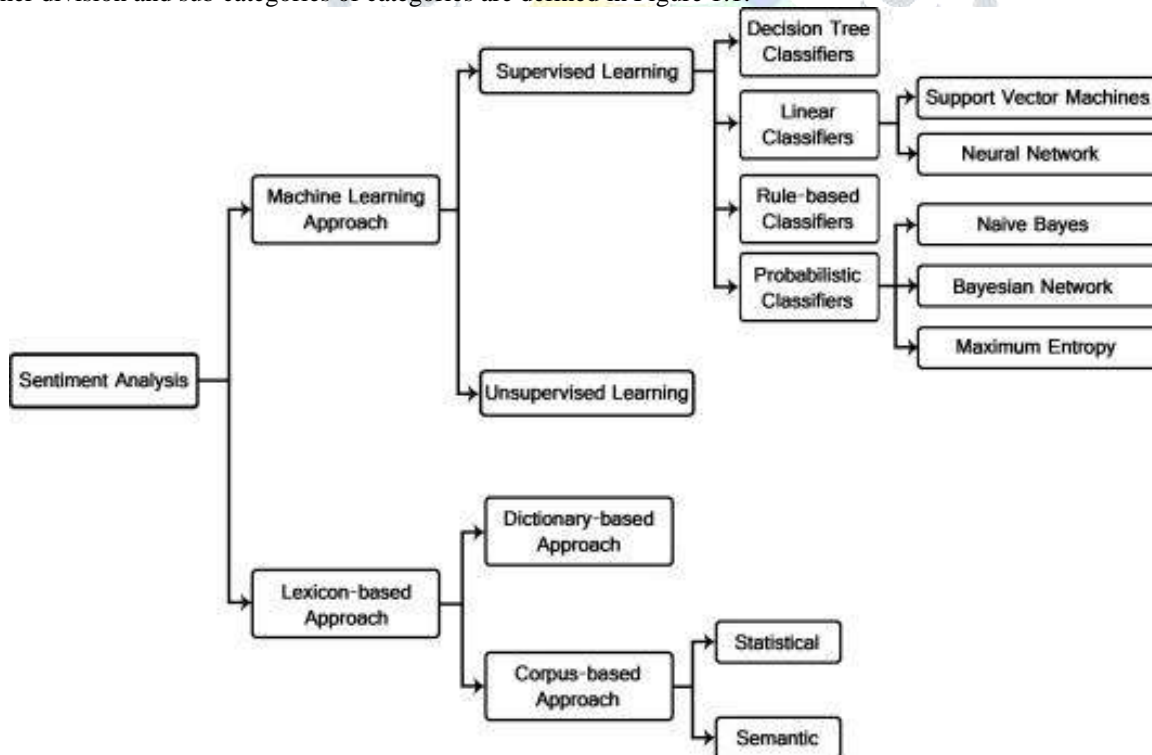

Figure 1.1 – Sentiment Analysis Techniques[12]

NAÏVE BAYES

It is very simple but a very powerful algorithm used in machine learning. This method is made up of two different types of probabilities which can be derived from the training data:

- o Probability of every class.
- o Conditional Probability of every class.

Calculated once, this method has the ability to get the predictions for newest data using Naïve Bayes theorem. If the data is real valued, then its common and easy to say that it's a Gaussian distribution to ease the estimation of the probabilities. Figure 1.5 below shows the Naïve Bayes method illustration:
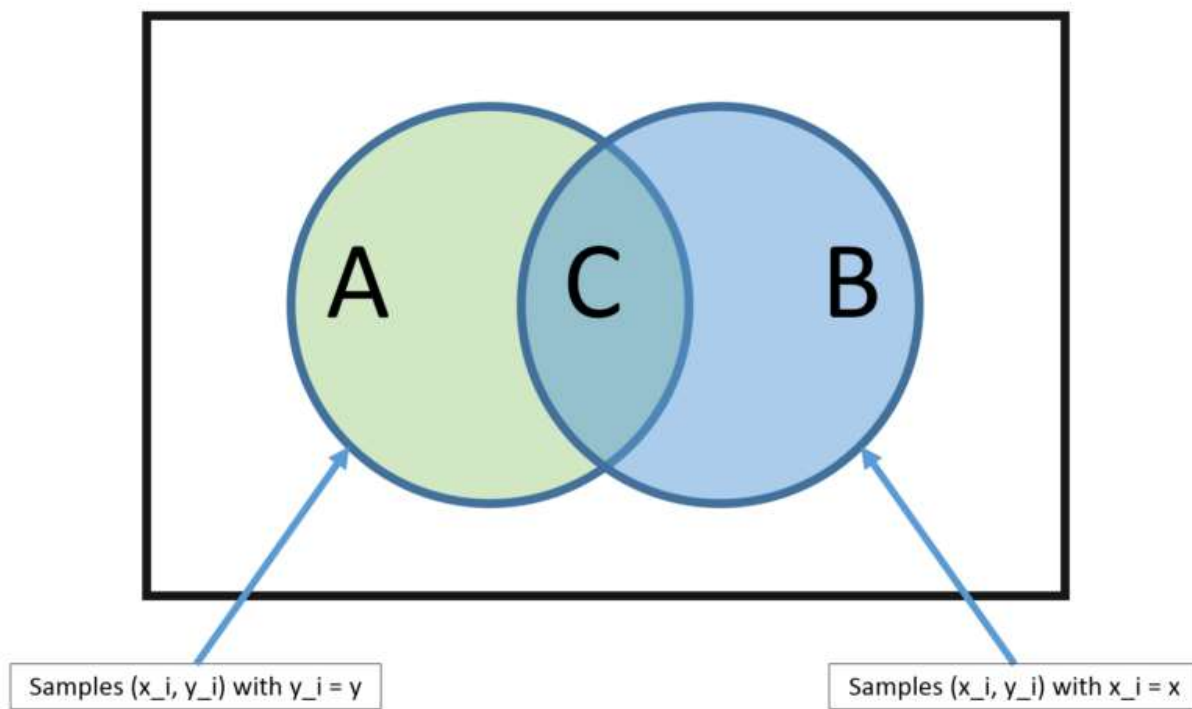

Figure 1.2 – Naïve Bayes [13]

It is known as naïve as it thinks that every input variable is totally independent. It is a very strong assumption but looks unrealistic for real time data. This algorithm works great for large number of issues. Naïve Bayes is a pretty simple and powerful techniques or algorithm for predictive modeling. When we are working with Machine Learning, we, most of the times are interested in getting the best hypothesis(h) given data(d) and one of the easiest methods or techniques of getting the hypothesis given the data which we can use or we already have of the problem. Naïve Bayes method is based on the Bayes theorem and is mainly used in cases where dimensionality of inputs are very high. It is simple, but very powerful and has the power to outperform other classification methods. We can use an example demonstration of NaiveBayes Classifier:
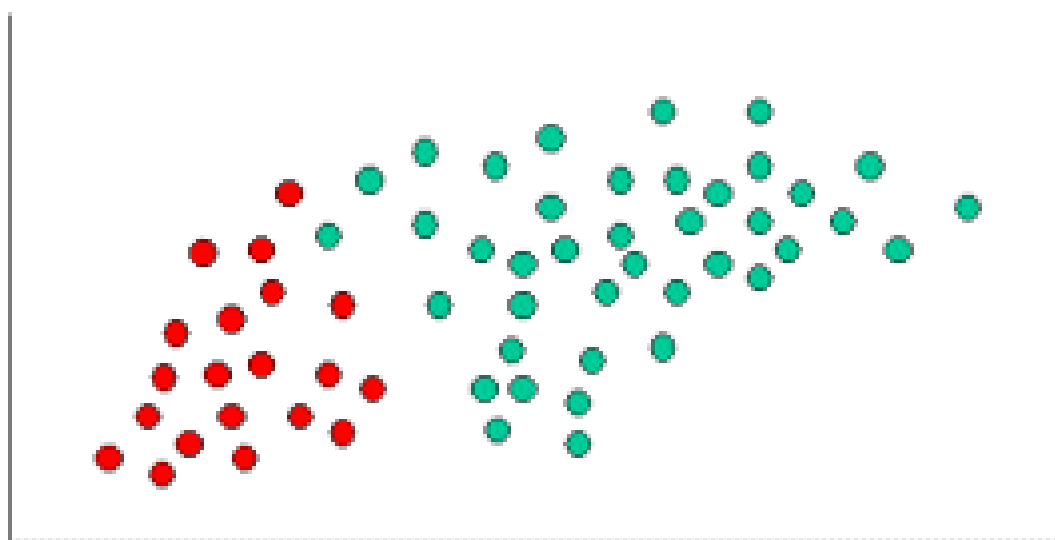

Figure 1.3 – Naïve Bayes Example

In the above illustration, objects can be seen as either red or green. Our work is to classify any new cases as they arrive and then decide on their label(red or green) on the basis of current object and their labels they have.

As seen green objects are almost twice the number as the reds, therefore it is reasonable to believe that a new case with no observation yet is twice likely to have the membership of green rather than red. As according to Bayesian, this is known as prior probability which is totally based on previous experience. We can also write the probability for Green and Red as:

$$Prior\ probability\ for\ GREEN \propto \frac{Number\ of\ GREEN\ objects}{Total\ number\ of\ objects}$$

$$Prior\ probability\ for\ RED \propto \frac{Number\ of\ RED\ objects}{Total\ number\ of\ objects}$$

Figure 1.4 – Prior Probability case

CONCLUSION

Sentiment Analysis is growing at a rapid pace in various industries for better customer experience and business productivity. Machine Learning and Deep Learning algorithms are used to perform text mining and to get the meaning information out of the structured or unstructured text. This bring better insight to the customers in the form of opinions, ratings or reviews etc. Different industries like Sports, Music, Movies, Politics, ecommerce etc. industries are using sentiment analysis and uses social network sites like facebook and twitter heavily to bring customer emotions and expressions on personality, product, organization pages, their posts, tweets to get to know about what people think. There are various levels and methodologies which can be used with the sentiment analysis like lexical affinity, keyword spotting and statistical method. Naïve Bayes is one of the most powerful algorithm to predict the sentiments from text mining.

REFERENCES

[1] **Anna Baj-Rogowska(2017),” Sentiment Analysis of Facebook Posts:the Uber case”,** IEEE International Conference on Intelligent Computing and Information Systems (ICICIS).

[2] **Antonio Teixeira and Raul M.S. Laureano(2013),”**Data extraction and preparation to perform the sentiment analysis using open source tools”.

[3] **Sanjida Akter and Muhammad Tareq Aziz(2016**),” Sentiment Analysis On Facebook Group Using Lexicon Based Approach”,IEEE.

[4] **Saud Alashri, Srinivasa Srivatsav Kandala, Vikash Bajaj, Roopek Ravi, Kendra L. Smith and Kevin C. Desouza(2016**),” An Analysis of Sentiments on Facebook during the

2016 U.S. Presidential Election”, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

[5] **N. AZMINA M. ZAMANI, SITI Z. Z. ABIDIN, NASIROH OMAR, M. Z. Z.ABIDEN(2013),”**Sentiment Analysis: Determining People's Emotions in Facebook”, Universiti Teknologi MARA, Malaysia.

[6]**Vishal A. Kharde and S.S. Sonawane(2016),”** Sentiment Analysis of Twitter Data: A Survey of Techniques”, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11.

[7] **Shufeng Xiong, Hailian Lv , Weiting Zhao, Donghong Ji(2017**),” Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings”, Elsevier.

[8] **Nehal Mamgain, Ekta Mehta, Ankush Mittal and Gaurav Bhatt(2016**),” Sentiment Analysis of Top Colleges in India Using Twitter Data”, International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT).

[9] **Aliaksei Severyn abd Aliaksei Severyn(2015),”** Twitter Sentiment Analysis with

Deep Convolutional Neural Networks”, ACM. ISBN 978-1-4503-3621-5/15/08.

[10] **Jˆonatas Wehrmann, Willian Becker, Henry E. L. Cagnini, and Rodrigo C. Barros(2017)**,” A Character-based Convolutional Neural Network for Language-Agnostic Twitter Sentiment Analysis”,IEEE.

[11] **Nhan Cach Dang, Fernando De la Prieta, Juan Manuel Corchado and María N. Moreno(2016**),” Framework for Retrieving Relevant Contents Related to Fashion from Online Social Network Data, Springer.

[12] **Walaa Medhat, AhmedHassan, Hoda Korashy(2014),** *“Sentiment analysis algorithms and applications: A survey” Ain Shams Engineering Journal.*

[13] **Towards Data Science, “**Top 10 algorithms for machine learning”. Retrieved from - https://towardsdatascience.com/a-tour-of-the-top-10-algorithms-for-machine-learning-newbies-dde4edffae11

[14] **T.M. Mitchel,** The Discipline of Machine Learning, CMU-ML-06-108, 2006

[15] **N. Cristianini and J. Shawe-Taylor.** An Introduction to Support Vector Machines. Cambridge University Press, 2000.