

A Comparative Study on Sentiment Analysis and Opinion Mining

Praveen Kumar¹, Umesh Chandra Jaiswal²

¹M.Tech Scholar, Department of Computer Science and Engineering,
MMMUT, Gorakhpur, INDIA
pbr162@gmail.com

²Associate Professor, Department of Computer Science and Engineering,
MMMUT, Gorakhpur, INDIA
ucj_jaiswal@yahoo.com

Abstract- In today's world opinions play important role in every aspects of life so, that it is an emerging topic for discussion. It can be calculated as sentimental analysis or opinion mining. All business and important events are based on analysis and feedback of customer reviews in a positive way therefore they increases productivity and performance related to services of items. In our paper we have expressed importance of sentimental analysis and various approaches to achieve bag-of –words in convenient and easy way using opinions techniques and with briefly represented by a diagram. In this paper we have illustrate important field of sentiment analysis for emotion detection, building resources, transfer learning etc. In paper we also covered different view of sentimental analysis approach, their comparison with showing table and in last we have compared different sentimental analysis approach.

Keywords- Sentiment analysis, semantics, lexicons, Natural Language Processing.

I. INTRODUCTION

Today we have huge amount of information that is available at different locations so, it is very difficult task to retrieve information from there, which causes various problems. So we have tried to demonstrate some rules guidelines and techniques for ensuring reliability and convenient way of accessing information through websites using sentiment analysis or opinion mining which is based on feedback of customers in positive or negative responses so we have improved our products according to their needs. In recently there are various way through which we express our opinion continuous increasing on world wide web their views with the help of movies, car etc. extra classification algorithm play important role in categorized new document which is extracted from web opinion mining in a natural language processing and information extraction task in order to fulfill the emotion express in positive and negative comment by request, question and by analyzing large number of document.

The study of opinion mining can be performed of document based where as in sentiment the complete document summarize as negative, positive or objective. Opinion mining based on face while sentiment is classified on the basis of polarity. In section 1 we describes and needs of sentimental analysis or opinion mining techniques, in section 2 literature survey. In section 3 describe important field of sentiment analysis. In section 4 we describe techniques and various rules through which we achieve above properties. In section 5 comparison of sentiment analysis techniques, and in section 6 conclusions and future work.

II. LITERATURE RIVIEW

Kanakaraj et al.[1]have proposed methods for enhancing the sentiment analysis classification by using some semantic feature vectors and other by sample process adding context , since adding identity and semantic similar word to feature vectors will increase accuracy of production because sentiment analysis system required bag of word for mining sentiment techniques for online reviews and social media .So using concept of semantic base feature vectors and together with ensemble classifier improved traditional bag of word techniques in a single machine learning classifier up to 3-5 percentage.

Gao et al. [2] have explore three web services which have different functionalities regarding sentimental analysis and compares their performances when they are classified text in tourism where all have right to public access and easy to use. In paper authors collected hotel review from trip adviser basically for collection for experiment and found that web services have high accuracy.

Kalarikkal et al. [3] has drawn attention towards different sentiment analysis method and role of data set for estimation for example SVM, NB, Maximum Entropy method. By using good quality training set we can increase efficiency of training set of sentiment analysis for better performance, so we used N-gram evaluation in case of word by word analysis. We know that in social networking site we get all important information by improving reviews and blogs for performance of website.

Aravindan et al. [4] have focus on importance of web application because they provided convenient way to buy product in their home easily which takes review from customers in order to increase productivity and

performance of services. There are various reviews which define various feature and product of services, customer go through various stages of reviews so they are fully decide to buy a product or which help in reviews for finding automatically extract relevant information of product and they can be reviews positive and negative reviews from the product feedback. The algorithm works on two stages first feature extraction and polarity classification. In this we applied association rule mining for knowing important properties of product. In second step we enhance supervised machine learning algorithm which is based on polarity classifier which determines sentiment analysis reviews of sentences on the basis of prominent feature so, they increases efficiency and achieve accuracy up to 79.67 percentages.

Ganeshbhai et al. [5] have focused on publish information over web which is generally in terms of comments and reviews they mainly toward analyzing the content towards visualization and public opinion. Opinion mining is self-govern text analysis and summarizations of things available on webs which controls our emotion and distinguish positive and negative aspect for analyzing positive and negative emotion of user.

Jha et al. [11] has proposed a new technique called Hindi Opinion Mining System which is efficiently using in review on movies data. We know that in the development of web technology field they are various documents which result different issue in opinion method. Online posting review is a way today in which we can express our opining and sentiment to words in trusted product for use as well services. On the basis of analysis large volume of opinion data are available on the web, therefore it is easy way to identified values of vendor and interested party. In today world web content document drastically increasing so there are need of execution of opinion mining a different languages, authors expressing different functionalities of useful a valuable Hindi Opinion Mining System which is generally perform task of their opinion mining on the basis of document level and classifier the document at three stage of level known as positive, negative and neutral by utilizing Port-Of-Speech tagging and Machine Learning technique, to perform our technique the use Naïve Bayes classifier.

Mishra et al. [12] has classified different types of opinion mining technique which can be used for remembering classifying establishing opinion orientation of the opinionate text. The main task of opinion mining to conclude negative or positive opinion document and sentences, author observe that in compound sentence it is tedious job for opinion mining. Various task of opinion mining are (1): Task of opinion mining document level. (2): Task of opinion mining sentence level. (3): Task of opinion mining feature level.

III. IMPORTANT FIELD OF SENTIMENT ANALYSIS

A. Emotion Detection

In some cases sentiment analysis perform NLP task for discovering opinion about an item this is happens because there is ambiguities in different type of opinion sentiment and emotion. Emotion detection is a kind SA task in which SA is mainly used for positive and negative opinion but ED is identified and detecting from various text. It can be implemented by using machine learning and lexicon based approach, generally used a lexicon based approach.

B. Building Resources

Its main task to creating lexicon, dictionary and corpora in which expression are major according to the polarity, building resources is a not a task SA but it can improve ED as well SA. The important challenges for this category are multi granularity ambiguities of word and difference in their opinion in text and gesture.

C. Transfer Learning

It extract knowledge for assisting domain for improving the learning process in target domain for example, it transfers knowledge for Wikipedia document for searching English to Arabic it use for enhance text mining task like sentimental analysis, text classification, part-of-speech tagging name entity recognition. In sentiment analysis this learning is applied some time for transfer sentimental classification in one domain to another.

IV. SENTIMENT ANALYSIS APPROACH

A. Machine Learning Approach

The machine learning approach is further useful in opinion mining than other approach due to fully automatic completion and its capability to hold large collection of Web Data. Machine Learning-Based Sentiment classification techniques can be classified into three types: supervised, unsupervised, and semi-supervised learning techniques.

- 1) *Supervised learning*: Supervised learning is our prime and successful key in traditional topical classification and has been adopted and investigated for opinion detection with satisfactory outcome. Supervised classification techniques are: a) Linear classifiers given $\bar{X} = \{x_1 \dots x_n\}$ is the normalized document word frequency, the output of linear classifier $p = \bar{A} \cdot \bar{X} + b$ is obtained by vector $\bar{A} = \{a_1 \dots a_n\}$ is a vector coefficients with the similar dimensionality as the attribute space. These are most commonly linear classifier are: Support Vector Machines (SVM), the main objective of (SVM) is to find out linear separators which suited for separating different classes without prior assumptions. b) Probabilistic classifier use

combination models for classification. These are most commonly probabilistic classifier are: a) Naïve Bayes classifier (NB) is convenient to use it compute the posterior probabilities of class which is based on distribution of word in document. It apply to predict the probability that use a given attribute

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

Here P(label)the prior probability of a label or the likelihood that a random feature set the label. P(features|lebal) is the prior probability that a given feature set is being classified is a lebal. P(features) is a prior probability that a given feature set is occurred [10]. b) Bayesian Network (BN) is the main assumption independence of feature and they are fully dependent of other features. This lead to a directed cyclic graph where nodes represent random variable as a conditional dependency. c) Maximum Entropy classifier(ME) a convert known feature set to vectors by imposing coding which is use for calculating weight and feature which can together combine and determine most common feature for a set.

2) *Unsupervised Learning*: It is easy to collect labeled and unlabeled document with help of unsupervised learning LDA and pLSA are examples of unsupervised learning thus advantage of unsupervised learning is to large amount of data for accurately training purpose.

3) *Semi-Supervised learning (SSL)*:It’s learned from label data examples SSL which learn label and unlabeled data and is new machine learning techniques in opinion mining. SSL is most famous algorithm which is used in self-training, generative models, co-training, multi-view learning, and graph based methods.

B. Lexicon-Based Approach

The lexicon-based approach based on sentiment lexicon which is integration of known and precompiled sentimental terminology, this is mainly classified dictionary-based approach and corpus-based approach. This approach based on to obtained opinion lexicon which are used in analyzing the text.

1) *Dictionary-based approach*: It depends on finding opinion and then searches dictionary of synonyms and antonyms. It has prebuilt dictionary that defines semantic orientation of word such as SentiWordNet which is standard dictionary today [1],[2],[3],[4],[5].

2) *Corpus-based approach*: It begin with seed list of opinion words than after find opinion of other word large corpus which help in finding opinion words in context orientation which can help of Statistical and Semantic approach.

a) *Statistical approach*: is for obtained polarity values be corpus which is specific dictionary creation to achieve adoptability to solve problem of unavailability of word uses by large corpus.

b) *Semantic approach*: is a based on a various principle calculating the similarities between show its provide sentiment values directly.

C. Hybrid based Approach

It is a combination of machine learning and lexicon-based approach and is frequently use with sentimental lexicon express major role in different important method.

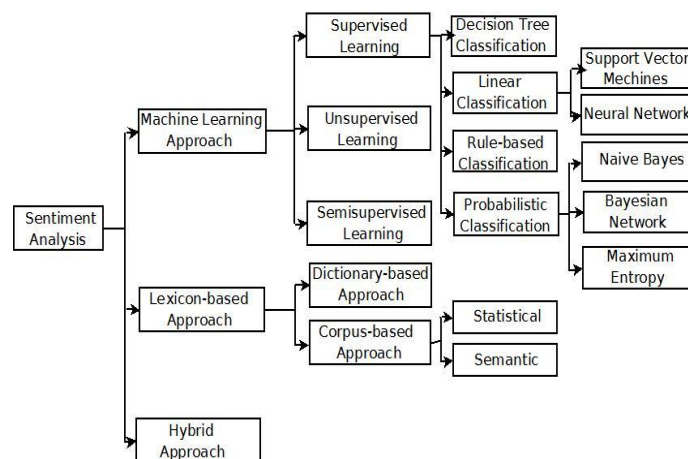


Figure.1. Sentiment Analysis approach

V. COMPARISSION OF DIFFERENT OPINION MINING TECHNIQUE ON THE BASIS OF VARIOUS LEVELS

In table1 authors have compare a two major approaches of sentiment analysis in which first is the Dictionary based techniques refers to a pre-built dictionary used for classifying the sentiment. The strength of classification in dictionary based depends on the reference dictionary used mainly these systems use Bag-of-Words thought which requires domain/context based semantics. Second is machine learning based which achieves domain based sentiment classification due to presence of domain specific training classification methods [1],[7],[8],[9].

Table.1. Comparison of existing approaches.

	Dictionary Based	Machine Learning Based
Approach	Uses references dictionary for classifying sentences/words Individual	Uses Supervised or Unsupervised ML algorithms for probability based classifications
Advantage	Lesser computation overhead as no training and modelling required	Can be customized to work more suitably for specific domains
Limitation	Lacks context or domain based classification capability	Affected by class imbalance and linguistic variations problems

We know that opinion mining technique at sentence level consist of single opinion posted by single opinion holder which is not suitable for multiple opinion in complex and compound sentences, it also have bond limitation of every document. The main task of this technique is to judging subjected and opinionated properties of sentences, it consist of positive, negative and neutral level.

In opinion mining a document level constricted on signal object and consists of opinion which is also posted by single opinion holder, but it is not appropriate for blog and forum post because there multiple opinions at various levels it is not easy to handle this entire task a single resource. It also consists of positive, negative and neutral level.

In opinion mining a feature level data search focus on single opinion holder which is posted by single object. It can be only applicable in single forum and blog and not applicable for multiple blogs and post due to there are different number of opinion for visualizing different object it's become tedious job to handle it. It determine whether the opinion feature of object are positive, negative and neutral in group feature it result feature based opinion for multiple review. And it last but not least it also extracted object opinion and feature based on reviewer comment.

Table.2. Comparison of various Sentiment Analysis methods.

S.no	Studies	SA Technique	Feature selection	Data Source	Performance	Precision	Recall
1	Pang and Lee(2004)	Naive bayes	Based on minimum	Movie review	86.4%	-	-
2	Hu and Liu (2005)	Opinion word	Opinion word	AmazonC mn.net	DVD-73%	-	-
3	Konig (2006)	hybrid	n-gram	Movie review	91%	-	-
4	Godbole (2007)	Lexical Approach	Graph distance	Blogs posts	82.7%-95.7%	-	-

5	Songho tan (2008)	Centroid classifier,	ML, G, CH	Chnsentic orp	90% SVM	-
6	Melville (2009)	Bayesian classifier,	n-gram	blogs	Blog 91.21%	-
7	Rudy (2009)	SVM, hybrid	Document	Movie review	89	-
8	Gangam (2010)	Maximum entropy	Dependency relation	Amazon review	-	78.7%
9	Kaigua n (2011)	Multiclass SVM	Linguistic Feature	Amazon Review	61%	93.4%
10	M.Taylor (2013)	Association rule	Holistic Linguist	Hotels Review	-	93%
11	D Arora and Kin	Lexical approach	Opinion word	Twitter API Twitte	95%	-

VI. CONCLUSION

In this paper we have illustrated importance of sentiment analysis and opinion mining in today's era for fetching relevant information exists from various sites on the basis of their opinion, feedback provided by customers and describes various techniques to achieve it. In future we will improve percentage of feedbacks from customer by providing opinions and various rules and results generated from opinion mining techniques. In this paper we saw that important field of sentimental analysis which perform various function on the basis of the functionalities in future we will implemented a such type of system which have ability of opinion mining and sentimental analysis according to our needs and categorized them on the basis of their classification and needs so therefore we will achieve performance, time and cost also. Through this we achieve our desirable things according to our needs. In future in dictionary based method we will solve context or domain based classification capability so therefore it reduces space and time complexity of new system this cause revolution in field of Artificial Intelligence. In future we will also solve problem in machine learning based problems by solving affected class imbalance problem and linguistic variations problems by trained our classes and their abstraction starts on the basis of their clusters and untrained classes are re trained so we get some trained classes from also their which leads to better utilization with efficient way and improves important features of system.

REFERENCES

- [1] M. Kanakaraj, R. Mohana, and R. Guddeti, "NLP Based Sentiment Analysis on Twitter Data Using Ensemble Classifiers," in 3rd International Conference on Signal Processing, Communication and Networking(ICSCN), 2015.
- [2] S. Gao, J Hao, Y Fu, "The Application and Comparison of Web Services for Sentiment Analysis in Tourism."
- [3] S. Kalarikkal, Remya P.C. "Sentiment Analysis and Dataset Collection : A Comparitive Study," 2015 IEEE International Advance Computing Conference(IACC),pp. 519–524, 2015.
- [4] S. Aravindan, A. Ekbal, "Feature Extraction and Opinion Mining in Online Product Reviews,"2014 IEEE International Conference on Information Technology, 2014.
- [5] S. Y. Ganeshbhai, Bhumika K. Shah, "Feature Based Opinion Mining : A Survey," 2015 IEEE International Advance Computing Conference (IACC) pp. 919–923, 2015.
- [6] B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technology , pp.1-167, 2012.
- [7] C. S. Analysis, "New Avenues in Opinion Mining and Sentiment Analysis," no. April, pp. 15–21, 2013.
- [8] N. R. Kasture, "An Approach for sentiment analysis on social networking sites," 2015 International Conference on Computing Communication Control and Automation(ICCCCA),2015.
- [9] L. Lin, J. Li, R. Zhang, W. Yu, and C. Sun, "Opinion Mining and Sentiment Analysis in Social Networks : A Retweeting Structure-aware Approach," 2014.
- [10] W. Medhat, A. Hassan, H. Korashy, "Sentiment analysis algorithms and application: A survey" Science Direct Journals,pp. 1093-1113, 2014.

- [11] V. Jha, N. Manjunath, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, "HOMS : Hindi Opinion Mining System," pp. 366–371, 2015.
- [12] N. Mishra and C.K.Jha, "Classification of Opinion Mining Techniques," *Int. J. Comput. Appl.*, vol. 56, no. 13, pp. 1–6, 2012.
- [13] D. Arora, K. F. Li, and S. W. Neville, "Consumers' Sentiment Analysis of Popular Phone Brands and Operating System Preference Using Twitter Data: A Feasibility Study," *2015 IEEE 29th Int. Conf. Adv. Inf. Netw. Appl.*, pp. 680–686, 2015.
- [14] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment Analysis of Short Informal Texts," *J. Artif. Intell. Res.*, vol. 50, pp. 1–15, 2014.
- [15] J. Cao, K. Zeng, H. Wang, and J. Cheng, "Web-Based Traffic Sentiment Analysis : Methods and Applications," vol. 15, no. 2, pp. 844–853, 2014.
- [16] S. Aravindan and A. Ekbal, "Feature Extraction and Opinion Mining in Online Product Reviews," *2014 Int. Conf. Inf. Technol.*, pp. 94–99, 2014.
- [17] A. Cernian, V. Sgarciu, and B. Martin, "Sentiment analysis from product reviews using SentiWordNet as lexical resource," pp. 2–5, 2015.
- [18] W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, "Sentence Compression for Aspect-Based Sentiment Analysis," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 13, pp. 2111–2124, 2015.
- [19] S. C. Harris, L. Zheng, V. Kumar, and Kinshuk, "Multi-dimensional Sentiment Classification in Online Learning Environment," *2014 IEEE Sixth Int. Conf. Technol. Educ.*, pp. 172–175, 2014.
- [20] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment Analysis of Short Informal Texts," *J. Artif. Intell. Res.*, vol. 50, pp. 1–15, 2014.
- [21] J. Cao, K. Zeng, H. Wang, and J. Cheng, "Web-Based Traffic Sentiment Analysis : Methods and Applications," vol. 15, no. 2, pp. 844–853, 2014.
- [22] S. Aravindan and A. Ekbal, "Feature Extraction and Opinion Mining in Online Product Reviews," *2014 Int. Conf. Inf. Technol.*, pp. 94–99, 2014.
- [23] A. Cernian, V. Sgarciu, and B. Martin, "Sentiment analysis from product reviews using SentiWordNet as lexical resource," pp. 2–5, 2015.