THINK INDIA JOURNAL
ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

Analysis of Classification Based Sentiment Analysis Techniques

Dr. Rajni Aron[1], Jyoti Godara[2]

1, Associate Professor, Lovely Professional University, Punjab

2, Assistant Professor, Lovely Professional University, Punjab

Abstract

The sentiment analysis is the major issue of the prediction analysis. The sentiment analysis has the various like data set collection, feature extraction and classification. The dataset is collected from the twitter using the twippy API. The feature extraction is the second phase in which relationship between the target set and tweets are established. In the last phase, classification phase will be applied which can classify data into certain phases positive, negative and neural phases. The various exists techniques for sentiment analysis for reviewed in terms of certain parameters

Keywords

Classification, Sentiment analysis, Feature extraction

Introduction

A kind of normal communication dispensation is known as sentiment or emotional evaluation to know the customer's opinion about a meticulous item. The other term of the study of feelings is mending perceptionsorbeliefs.This analysis develops a scheme for collecting and examining views on a specific o bject in social media posts, evaluations, tweets, or remarks The approach to emotional analysis can be us eful in various ways.This research illustrates its influence in government and business, ranging from co mputer discipline to organizational and public science discipline.

In recent years, emotion-related manufacturing operationshavealsoflourished.Different novel industries were developed.Most big businesses are selfreli ant on domestic capacity. Emotional screening systems have made their claims in roughly every industry and public area

[1].Opinion analysis can be described as a procedure in the form of content, language, chirp, record, etc. involving the extraction, estimate, vision and feeling of computerized feelings from the processing of nat ural language. The tweets were mostly grouped into three groups during the Opinion investigation.These are "optimist," "pessimistic" and "unbiased" categories. This research can also be int erpreted in the context of bias inquiry, beliefs extraction, and assessment of mining.Sarcasm is a form of verbal irony in which contempt or ridicule is expressed.Sarcasm has an implied negative feeling, but may not have a negative feeling on the surface[2]. A sarcastic sentence can have a positive feeling on the surface, a negative feeling on the surface or no feeling on the surface. Sarcasm therefore implies feeling. Sarcasm is one of the major challenges faced in the analysis of sentiments. Detection of sarcasm is important in sentiment analysis because such expressions can flip the polarity of a sentence. Presence of sarcasm makes the task of understanding the opinions of users and performing sentiment analysis

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

very difficult, therefore the detection of sarcasm is important. Approaches for detection of sarcasms were categorized into a rule-based approach, a statistical approach, and a deep learning-based approach. A rule-based system uses semant and syntactic rules to analyze and generate[3].
Classification of emotions can be achieved in two ways: machine learning approach and methods focused on lexicons.Machine learning technique engages both supervised and unsupervised learning. Unsupervised Learning approach is relied on merely key, devoid of any target mentioning. This technique mainly relies on clustering. It does not contain any category. Also the accurate objectives are not assessable to this techniques therefore they just depend on clustering. Supervised Learning approach describes earlier précised objectives and the attainment of these objectives is necessary in association with inputs. Information samples are trained for the attainment of important harvest while stumble upon in the period of execution. Lexicon-Based approach allocates optimistic or pessimistic division relied on the opinion of every remark and on the basis of this, a vocabulary is shaped [4]. With the help of a unite function like sum or average, the universal outlook of a paper can be find out. This methodology utilizes emotion vocabulary in association with estimation terms and matches them with information for the determination of division. They give opinion grades to the estimation texts for the description of optimistic, pessimistic and unbiased texts. Lexicon relied advances mostly depends upon a emotion lexicon i.e., compilation of recognized and earlier accumulated emotion stipulations, sentences and even idioms, structured for conventional varieties of interaction in the form of estimation locating lexicon. Hybrid method merges both machine learning and lexicon relied techniques for the attainment of improved categorization outcomes [5]. The prevailing characteristics of these approaches may be implemented for the attainment of balanced outcomes. Although hash tag-based labeling may provide comprehensive monitoring, the database accuracy may be uncertain. In the case of using #it is especially true not to suggest a sense of insincerity. In addition, hash tag (#) is not often used to express sarcasm while the remainder of the sentence is not sufficient to identify the sarcasm.However, it should be noted that these approaches require ' surface polarity ': a sentence's apparent polarity[6]. Thus, when sentiment is used as a feature, several issues arise. Sarcasm is a rare phenomenon of expression of feelings. For databases, this bias also represents. Dealing with skewed datasets is not easy.

Literature Review

PornimaGidhe, et.al (2017) stated that a lot of researches were conducted on dataset including # sarcasm tag. These researchers mostly used structural and sentiment attributes. However, some cases did not include # sarcasm tags [7]. Emotional and semantic similarity features were needed to transcribe these types of cases. In this work, a novel scheme had been recommended for the classification of sarcastic and non-sarcastic sentences. For this purpose, structural, emotional and semantic similarity features with MLP-BP approach were used in this work.

Dwi A. P. Rahayu, et.al (2018) Suggested a new approach to detection of sarcasm in Indonesian tweets. Two feature extraction techniques were used by the proposed approach [8]. These techniques were called interjection and punctuation. Afterward, two different weighting and classification algorithms used these techniques. The experimental outcomes depicted that optimum performance had been shown

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

by the proposed approach in sarcasm detection by combining feature extraction techniques i.e. tf-idf and k-Nearest Neighbor.

Muhammad Abulaish, et.al (2018) suggested a new technique for self-deprecating sarcasm detection. The proposed approach combines sarcasm detection algorithms based on rules and machine learning.[9].
Three different classification models were trained by defining 11 attributes, named decision tree, naïve Bayes, and bagging. Such features included six self-deprecating characteristics and five hyperbolics. A Twitter dataset having 107536 tweets was used to evaluate the recommended scheme. In this work, Comparison of proposed approach and some state-of-the-art techniques was carried out to detect sarcasm.

Paras Dharwal, et.al (2017) stated that it was a very complex task to detect sarcasm in automatic manner in Sentiment Analysis [10]. The detection task included composite linguistic analysis and machine learning algorithms. In this work, a lot of sarcasm analysis techniques were reviewed. These techniques were used to filter the sarcastic statements from a text. These techniques also analyzed the application of Automatic sarcasm detection in the classification of product review texts and tweets.

Anandkumar D. Dave, et.al (2016) made an attempt to recognize various supervised classification algorithms employed to detect sarcasm detection and its features [11]. In this work, results provided by different classification algorithms on existing textual data available on different social media sites in different languages were also analyzed.I nitially, some tests were conducted for detecting sarcastic sentences in "Hindi" text. It was identified that a simple model based on "bag-of-words" feature had the ability to classify half of the sarcastic sentences in accurate manner.

Md Saifullah Razali, et.al (2017) stated that it was possible to use sarcasm detection approach for improving sentiment analysis [12]. In contrast to text, only few researchers were conducted on sarcasm through multimodalities in sentiment analysis domain. The available research depicted the rising use of multimodalities, particularly from social media platform. It was advantageous to use massive volume of online available multimodal data for those researches which were merely based on text-only analysis.

Anukarsh G. Prasad, et.al (2017) compared different classification algorithms for detecting sarcasm within tweets [13]. These tweets were obtained from the Twitter Streaming API. The optimum classification was selected after comparative analysis. The selected classifier was combined with several pre-processing and filtering methods. For this purpose, emoji and slang dictionary mapping were used for getting the conceivable accuracy rate. In future, the achieved outcomes would be utilized as input to assist other research works.

Santosh Kumar Bharti, et.al (2017) stated that several sarcasm detection approaches had been recommended for detecting sarcastic within text [14]. Detecting sarcasm in English text was the main

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

aim of almost all techniques. A novel context-based pattern had been recommended in this work to detect sarcasm within Hindi tweets. In the proposed pattern, sarcasm was represented as a conflict between a tweet and the reference of its relevant news. The suggested technique used Hindi news as the reference of a tweet within the similar timestamp and achieved 87% accuracy rate.

Ravinder Ahuja, et.al (2018) implemented twelve different classification algorithms on four sorts of datasets [15]. These dataset types were called Set1, Set2, Set3 and Set4. In order to verify the accuracy of all algorithms in different conditions, the split ratio of the datasets was changed accordingly. In this work, behavioral approach had been implemented for detecting sarcasm on twitter dataset. However, gradient improvement was noted in the set 4 that provided optimal accuracy in all three split ratio cases. These cases were 50:50, 25:75, 10:90 and the achieved accuracy rate was 85.14%, 85.71%, and 85.03% respectively in all three cases.

Table 1:Table of Comparison

Author Name	Year	Technique used for sarcasm detection	Merits	Results	Demerits
Pornima Gidhe, et.al	2017	MLP-BP (Multilayer perceptron-Backpropagation)	The semantic similarity can be identified through this proposed approach.	In this work, a novel scheme had been recommended for the classification of sarcastic and non-sarcastic sentences. For this purpose, structural, emotional and semantic similarity features with MLP-BP approach were used in this work.	This study does not provide a comparative analysis of proposed and previously existing approaches.
Dwi A. P. Rahayu, et.al	2018	tf-idf and k-Nearest Neighbor	In comparison to the combined naïve bayes, interjection and punctuation based techniques; the proposed approach provided better outcomes.	The experimental outcomes depicted that optimum performance had been shown by the proposed approach in sarcasm detection by combining feature extraction techniques i.e. tf-idf and k-Nearest Neighbor	The research was limited to Indonesia only.
Muhammad et.al	2018	Decision tree, naïve Bayes, and decision tree bagging, naïve Bayes, and	This approach aimed to handle the self-deprecating	A Twitter dataset having 107536 tweets was used to evaluate the recommended scheme. In this work,	Other categories of sarcasm were not focused here, such as broo

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

		decision tree bagging, naïve Bayes, and rule-based strategies bagging.	sarcasm detection based problems.	Comparison of proposed approach and some state-of-the-art techniques was carried out to detect sarcasm.	ding, polite, raging, etc.
Paras Dharwal, et.al	2017	Support Vector Machines, Logistic Regression, Naïve Bayes, tf-idf	Accuracy was increased to great extent by merging the two mentioned approaches.	These techniques also analyzed the application of Automatic sarcasm detection in the classification of product review texts and tweets.	The use of Naïve bayes in this approach could lead to low precision value resulting is poor outcomes.
Anandkumar D. Dave,et. al	2016	TF-IDF, Naive Bayes Classifier, Maximum Entropy Classifier, Support Vector Machine, Conditional Random Field (CFR)	Twitter, Amazon and social media sites are primary and easily accessible as per this research.	It was identified that a simple model based on "bag-of-words" feature had the ability to classify half of the sarcastic sentences in accurate manner.	Sarcasm detectionalso depends on context of text available which was not discussed here.
Md Saifullah Razali, et.al	2017	rule-based approaches, pattern-based features, deep learning	There was an increase in use of multimodalities which was discussed through this research.	The available research depicted the rising use of multimodalities, particularly from social media platform. It was advantageous to use massive volume of online available multimodal data for those researches which were merely based on text-only analysis.	This research only focused on text-analysis based approaches.
Anukarsh G. Prasad, et.al	2017	Random Forest, Gradient Boosting, Decision Tree, Adaptive Boost, Logistic Regression and Gaussian Naïve Bayes	The real-time based scenarios were used to conduct experiments which provided a good emphasis on the outcomes.	Emoji and slang dictionary mapping were used for getting the conceivable accuracy rate.	The effectiveness of techniques was not discussed here.
Santosh	2017	context-based	Based on	The suggested technique	Only the Hindi

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

Kumar Bharti, et.al		pattern	similar timestamp the data was collected and results were evaluated.	used Hindi news as the reference of a tweet within the similar timestamp and achieved 87% accuracy rate.	new related data could be examined here.
Ravinder Ahuja, et.al	2018	Gradient Boosting, Gaussian Naive Bayes, Adaboost etc.	Highly robust results were achieved even with the partial implementation of the SCUBA framework.	Gradient improvement was noted in the set 4 that provided optimal accuracy in all three split ratio cases. These cases were 50:50, 25:75, 10:90 and the achieved accuracy rate was 85.14%, 85.71%, and 85.03% respectively in all three cases	The current approach was only examined on static dataset.

Conclusion

In this work, it is concluded that sentiment analysis is the issue of prediction analysis due to complex attributes. The sentiment analysis has the various phases like data set collection, features extraction and classification. In this paper, classification techniques for the sentiment analysis are reviewed in terms of certain parameters. In future, hybrid classification technique for sentiment analysis will be designed which give high accuracy

References

[1] Kiruthika M, Sanjana Woonna, “Sentiment analysis of twitter data”, 2016, International journal of innovations in engineering and technology.

[2] Varsha Sahayak, Vijaya Shete, Apashabi Pathan, “Sentiment Analysis on Twitter Data”, 2015, International Journal of Innovative Research in Advanced Engineering (IJIRAE)

[3] Apoorv Agarwal BoyiXie Ilia Vovsha Owen Rambow Rebecca Passonneau, “Sentiment Analysis of Twitter Data”, 2011, Conference of the European Chapter of the ACL

[4] Chandan Arora, Dr. Rachna, “SENTIMENT ANALYSIS ON TWITTER DATA”, 2017, International Research Journal of Engineering and Technology (IRJET)

[5] Vishal A. Kharde, S.S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques”, 2016, International Journal of Computer Applications (0975 – 8887)

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

[6] Onam Bharti, Mrs. Monika Malhotra, "SENTIMENT ANALYSISON TWITTER DATA", 2016, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.6,

[7] PornimaGidhe, Leena Ragha, "Sarcasm detection of non # tagged statements using MLP-BP", International Conference on Advances in Computing, Communication and Control (ICAC3), Year: 2017 | Conference Paper | Publisher: IEEE

[8] Dwi A. P. Rahayu, SoveatinKuntur, Nur Hayatin, "Sarcasm Detection on Indonesian Twitter Feeds", 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Year: 2018 | Conference Paper | Publisher: IEEE

[9] Muhammad Abulaish, Ashraf Kamal, "Self-Deprecating Sarcasm Detection: An Amalgamation of Rule-Based and Machine Learning Approach", IEEE/WIC/ACM International Conference on Web Intelligence (WI), Year: 2018 | Conference Paper | Publisher: IEE

[10] Paras Dharwal, Tanupriya Choudhury, Rajat Mittal, Praveen Kumar, "Automatic sarcasm detection using feature selection", 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Year: 2017 | Conference Paper | Publisher: IEEE

[11] Anandkumar D. Dave, Nikita P. Desai, "A comprehensive study of classification techniques for sarcasm detection on textual data", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Year: 2016 | Conference Paper | Publisher: IEEE

[12] Md Saifullah Razali, Alfian Abdul Halin, NorisMohdNorowi, Shyamala C. Doraisamy, "The importance of multimodality in sarcasm detection for sentiment analysis", IEEE 15th Student Conference on Research and Development (SCOReD), Year: 2017 | Conference Paper | Publisher: IEEE

[13] Anukarsh G. Prasad, S. Sanjana, Skanda M. Bhat, B. S. Harish, "Sentiment analysis for sarcasm detection on streaming short text data", 2nd International Conference on Knowledge Engineering and Applications (ICKEA), Year: 2017 | Conference Paper | Publisher: IEEE

[14] Santosh Kumar Bharti, Korra Sathya Babu, Rahul Raman, "Context-based Sarcasm Detection in Hindi Tweets", Ninth International Conference on Advances in Pattern Recognition (ICAPR), Year: 2017 | Conference Paper | Publisher: IEEE

[15] Ravinder Ahuja, Shantanu Bansal, Shuvam Prakash, Karthik Venkataraman, Alisha Banga, "Comparative Study of Different Sarcasm Detection Algorithms Based On Behavioral Approach", Procedia Computer Science, Volume 143, 2018, Pages 411-418