ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

Sarcasm and Irony Word Analysis in Social Network Using Deep Learning Algorithm

Dr.Kalaimani Shanmugam [1], Haritha.G [2]

Professor and Head, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam, Tamilnadu, India [1]

P.G. Student, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam, Tamilnadu, India [2]

ABSTRACT: In recent years, social network sites like Facebook, Instagram and Twitter has acquired extensive popularity and importance. Twitter is one of the largest social platforms where people express their opinions, feelings, views and real-time events such as live tweets etc. Twitter allows the users to register and then read and send messages which are known as tweets. Sarcasm is one of the major challenges faced in Sentiment Analysis. Twitter also enables the users to express their ideas and opinions with each other which enable the companies to know the public opinion on their products or services so that they can provide the real- time customer assistance. Designing efficient and robust algorithms for detection of sarcasm on Twitter is the exciting challenge in opinion mining field. Sarcasm means the person speaks the contradictory of what the individual means, expressing gloomy feelings applying positive words. It helps the retailers to know the opinions of the customers. Sarcasm is widely used in many social networking and micro-blogging websites where people invade others which makes problematic for the individuals to say what it means. In the existing systems, machine learning technique is used to detect these sarcastic tweets, it has a drawback as it cannot predict for continuous variables. In the proposed methodology Sentiment Analysis, Deep learning algorithm is used to detect sarcasm, positive and negative words on twitter. By using deep neural network, the tweets are categorized into sarcastic and non-sarcastic and positive, negative words. Sentiment Analysis is used to mine the opinions of customers to identify and extract information from the text. By using back propagation neural network, sarcastic statements can be easily classified and identified from twitter.

KEYWORDS: Sentiment Analysis; opinion mining; social network; Deep learning algorithm; sarcastic statements

I. INTRODUCTION

1.1 MACHINE LEARNING

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to learn automatically without human intervention and to improve from experience without being explicitly programmed.

1.2 DEEP LEARNING

Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled. It is also known as deep neural learning or deep neural network.

1.3 DATA MINING

Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining processes can be classified into two types: data preparation or data pre-processing and data mining. In fact, the first four processes, that are data cleaning, data integration, data selection and data transformation, are considered as data preparation processes. The last three

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

processes including data mining, pattern evaluation and knowledge representation are integrated into one process called data mining.

1.4 PROCESS IN DATA MINING

a) Data Cleaning

Data cleaning is the process where the data gets cleaned. Data in the real world is normally incomplete, noisy and inconsistent. The output of data cleaning process is adequately cleaned data.

b) Data Integration

Data integration is the process where data from different data sources are integrated into one. Data integration tries to reduce redundancy to the maximum possible level without affecting the reliability of data.

c) Data Selection

Data selection is the process where the data relevant to the analysis is retrieved from the database.

d) Data Transformation

Data transformation is the process of transforming and consolidating the data into different forms that are suitable for mining. Data transformation normally involves normalization, aggregation, generalization etc.

e) Data Mining

Data mining is the core process where a number of complex and intelligent methods are applied to extract patterns from data.

f) Pattern Evaluation

Pattern evaluation identifies the truly interesting patterns representing knowledge based on different types of interestingness measures.

g) Knowledge Representation

The information mined from the data needs to be presented to the user in an appealing way. Different knowledge representation and visualization techniques are applied to provide the output of data mining to the users.

II. RELATED WORK

"SARCASM DETECTION ON TWITTER: A BEHAVIORAL MODELING APPROACH"by ASHWIN RAJADESINGAN, 2015 the paper contributes that in recent years, social media sites such as Twitter have gained immense popularity and importance. These sites have evolved into large ecosystems where users express their ideas and opinions uninhibitedly. Companies leverage this unique ecosystem to tap into public opinion on their products or services and to provide real-time customer assistance. Sarcasm is a nuanced form of language in which individuals state the opposite of what is implied. With this intentional ambiguity, sarcasm detection has always been a challenging task, even for humans. Current approaches to automatic sarcasm detection rely primarily on lexical and linguistic cues. This paper aims to address the difficult task of sarcasm detection on Twitter by leveraging behavioral traits intrinsic to users expressing sarcasm. We identify such traits using the users past tweets. We employ theories from behavioral and psychological studies to construct a behavioral modeling framework tuned for detecting sarcasm. We evaluate our framework and demonstrate its efficiency in identifying sarcastic tweets.

"A PATTERN-BASED APPROACH FOR SARCASM DETECTION ON TWITTER" by MONDHER BOUAZIZI, 2016 the paper contributes that twitter became one of the biggest web destinations for people to express their opinions, share their thoughts and report real-time events, etc. Throughout the previous years, Twitter content continued to increase, thus constituting a typical example of the so-called big data. Sarcasm is a sophisticated form of irony widely used in social networks and micro blogging websites. Therefore, recognizing sarcastic statements can be very useful to improve automatic sentiment analysis of data collected from micro blogging websites or social networks. Sentiment Analysis refers to the identification and aggregation of attitudes and opinions expressed by Internet users toward a specific topic. In this paper, we propose a pattern-based approach to detect sarcasm on Twitter. We propose four sets of features that cover the different types of sarcasm we defined. In this work, we proposed a new method to detect sarcasm on Twitter. The proposed method makes use of the different components of the tweet. Our approach makes use of Part-of-Speech tags to extract patterns characterizing the level of sarcasm of tweets. The approach has

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

shown good results, though might have even better results if we use a bigger training set since the patterns we extracted from the current one might not cover all possible sarcastic patterns.

III. PROPOSED ALGORITHM

3.1 PROPOSED SYSTEM

Sarcasm detection has been modeled as a binary document classification task, with rich features being defined manually over input documents. Sarcasm detection is the task of correctly labeling the text as 'sarcastic' or 'non-sarcastic'. It is a challenging task owing to the lack of intonation and facial expressions in text. Nonetheless humans can still spot a sarcastic sentiment in the text and reason about what makes it so. Recognizing sarcasm in text is an important task for Natural Language processing to avoid misinterpretation of sarcastic statements as literal statements. Accuracy and robustness of NLP models are often affected by untruthful sentiments that are often of sarcastic nature. Thus, it is important to filter out noisy data from the training data inputs for various NLP related tasks. NLP is a branch of data science that consists of systematic processes for analyzing, understanding, and deriving information from the text data in a smart and efficient manner.

By utilizing NLP and its components, one can organize the massive chunks of text data, perform numerous automated tasks and solve a wide range of problems such as – automatic summarization, machine translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation etc. For this prediction task we are using Back propagation Neural Network. A neural Network is a machine learning algorithm to perform Classification and Regression related task. We are using logistic sigmoid as activation function for the neural network. By using this proposed method, the advantages are reducing Time Complexity, improved accuracy rate, reduce the false positive rate.

3.1.1 DEEP LEARNING PROCESS

It is a part of Machine learning algorithms. For feature extraction and transformation it uses many different layers of non-linear processing units. Deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. In this module, we can implement back propagation neural network algorithm to train the keywords with labels.

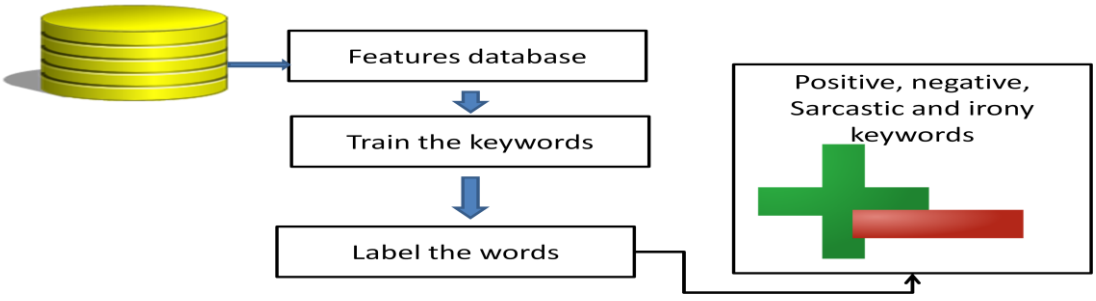

Fig.1. Deep Learning Process

The inputs from the input layer to the hidden layer are multiplied with the respective weights and then the each hidden node sums up all the inputs it is getting. Then the value is passed through the activation function and again the values from hidden layer to output layer are multiplied by respective weights and the output sums up the input it is receiving, then it passes the sum through the activation again and produces output. The output from the output layer is then compared with the target output. Our goal with back propagation is to update each of the weights in the network so that they cause the actual output to be closer the target output, thereby minimizing the error for each output neuron and the network as a whole.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

3.2 TEXT MINING ALGORITHM

Text mining, also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interestingness.

Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling (i.e., learning relations between named entities). The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods. A typical application is to scan a set of documents written in a natural language and either model the document set for predictive classification purposes or populate a database or search index with the information extracted. Text Analytics, also known as text mining, is the process of examining large collections of written resources to generate new information, and to transform the unstructured text into structured data for use in further analysis. Text mining identifies facts, relationships and assertions that would otherwise remain buried in the mass of textual big data. These facts are extracted and turned into structured data, for analysis, visualization (e.g. via html tables, mind maps, charts), integration with structured data in databases or warehouses, and further refinement using machine learning (ML) systems.

Text mining has become more practical for data scientists and other users due to the development of big data platforms and deep learning algorithms that can analyze massive sets of unstructured data. Mining and analyzing text helps organizations find potentially valuable business insights in corporate documents, customer emails, call center logs, verbatim survey comments, social network posts, medical records and other sources of text-based data. Increasingly, text mining capabilities are also being incorporated into AI chatbots and virtual agents that companies deploy to provide automated responses to customers as part of their marketing, sales and customer service operations. Mining and analyzing text helps organizations find potentially valuable business insights in corporate documents, customer emails, call center logs, verbatim survey comments, social network posts, medical records and other sources of text-based data. Increasingly, text mining capabilities are also being incorporated into AI chatbots and virtual agents that companies deploy to provide automated responses to customers as part of their marketing, sales and customer service operations.

Text mining is similar in nature to data mining, but with a focus on text instead of more structured forms of data. However, one of the first steps in the text mining process is to organize and structure the data in some fashion so it can be subjected to both qualitative and quantitative analysis. Doing so typically involves the use of natural language processing (NLP) technology, which applies computational linguistics principles to parse and interpret data sets. The upfront work includes categorizing, clustering and tagging text; summarizing data sets; creating taxonomies; and extracting information about things like word frequencies and relationships between data entities.

Text mining is the method of extracting meaningful information or knowledge or patterns from the available text documents from various sources. It contains following steps as follows

Step 1: Choosing the scope of document
Step 2: Tokenization
Step 3: Token Normalization
Step 4: Stop words removal
Step 5: Stemming the words
Step 6: Remove special characters

3.3 BACK PROPOGATION ALGORITHM

Back propagation algorithms are a family of methods used to efficiently train artificial neural networks (ANNs) following a gradient-based optimization algorithm that exploits the chain rule. The main feature of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

back propagation is its iterative, recursive and efficient method for calculating the weights updates to improve the network until it is able to perform the task for which it is being trained. It is closely related to the Gauss–Newton algorithm.

Back propagation requires the derivatives of activation functions to be known at network design time. Automatic differentiation is a technique that can automatically and analytically provide the derivatives to the training algorithm. To understand the mathematical derivation of the back propagation algorithm, it helps to first develop some intuition about the relationship between the actual output of a neuron and the correct output for a particular training example. Consider a simple neural network with two input units, one output unit and no hidden units, and in which each neuron uses a linear output (unlike most work on neural networks, in which mapping from inputs to outputs is non-linear)that is the weighted sum of its input.

In the context of learning, back propagation is commonly used by the gradient descent optimization algorithm to adjust the weight of neurons by calculating the gradient of the loss function; back propagation computes the gradient(s), whereas (stochastic) gradient descent uses the gradients for training the model (via optimization). BPP algorithms contains following steps as follows

Step 1: Randomly initialize the weights and biases.
Step 2: feed the training sample.
Step 3: Propagate the inputs forward; compute the net input and output of each unit in the hidden and output layers.
Step 4: back propagate the error to the hidden layer.
Step 5: update weights and biases to reflect the propagated errors.
Training and learning functions are mathematical procedures used to automatically adjust the network's weights and biases.
Step 6: terminating condition

IV. SIMULATION RESULTS

Thus the work is based on identifying the sarcastic statements from the tweets which is shown in the simulation result. The tweets can be classified into positive, negative, neutral, sarcasm. Thus the Fig.2 shows the match score.

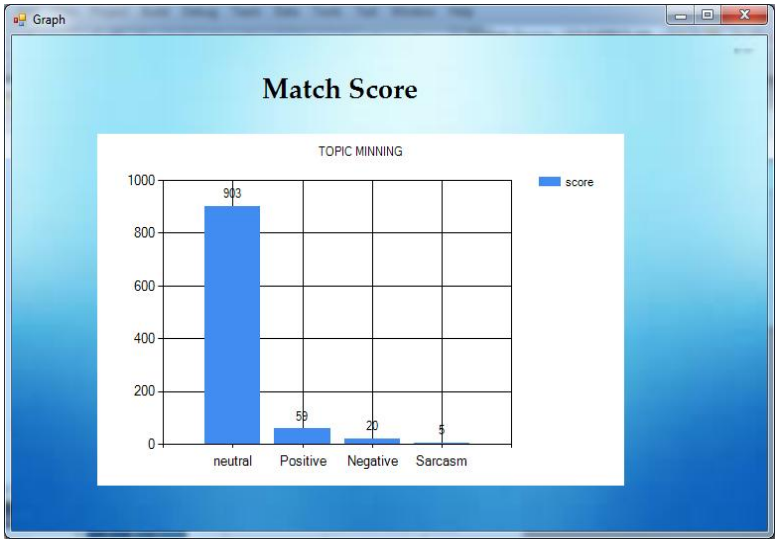

Fig.2. Match Score

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

V. CONCLUSION AND FUTURE WORK

Thus we conclude that the back propagation neural network technique is very easier and efficient than symbolic techniques. These techniques are easily applied to twitter sentiment analysis. Twitter sentiment analysis is difficult because it is very tough to identify emotional words form tweets and also due to the presence of the repeated characters, slang words, white spaces, misspellings etc. To handle these problems the feature vector is created. Before creating feature vector pre-processing is done on each tweet. Then features are extracted in two phases: First phase is the extraction of the twitter specific word. Then they are removed from the text. Now extracted feature vector is transformed into normal text. After that, features are extracted from tweet which is normal text without any hash tags or slang words. And these extracted features are then added to form feature vector. There are different machine learning classifiers to classify the tweet. It is proposed to stream real time live tweets from twitter using Twitter API, and the large volume of data makes the application suitable for Big Data Analytics. A method to predict or deduct the location of a tweet based on the tweet's information and the user's information should be found in the future.

In future work we can extend the framework to implement in real social network framework. And include unwanted word analysis, short terms analysis using deep learning process. Then will implement the framework with real time alert system at the time of blocking unwanted comments and also unwanted friends.

REFERENCES

[1]S. Lukin and M. Walker, "Really? well. apparently bootstrapping improves the performance of sarcasm and nastiness classifiers for online dialogue," in Proceedings of the Workshop on Language in Social Media (LASM), Atlanta, Georgia, pp. 30–40, ACL, June 13, 2013.
[2] E. Riloff, A. Qadir, P. Surve, L. D. Silva, N. Gilbert, and R. Huang, "Sarcasm as contrast between a positive sentiment and negative situation," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Seattle, Washington, USA, pp. 704–714, ACL, October 18-21, 2013
[3] M. Boia, B. Faltings, C.-C. Musat, and P. Pu, ``A :) Is worth a thousand words: How people attach sentiment to emoticons and words in tweets," in Proc. Int. Conf. Social Comput., Sep. 2013, pp. 345_350.
[4]T. Pt´acek, I. Habernal, and J. Hong, "Sarcasm detection on czech and english twitter," in Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland, pp. 213– 223, ACL, August 23-29, 2014.
[5]A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, "Semeval-2015 task 11: Sentiment analysis of figurative language in twitter," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval), Denver, Colorado, pp. 470– 478, ACL, June 4-5, 2015.
[6] A. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm detection on twitter: A behavioral modeling approach," in Proceedings of the 8th Association for Computing Machinery International Conference on Web Search and Data Mining (WSDM), Shanghai, China, pp. 97– 106, ACM, February 2-6, 2015.
[7]W. Gao and F. Sebastiani, ``Tweet sentiment: From classification to quantification," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2015, pp. 97_104.
[8] Y. H. P. P. Priyadarshana, K. I. H. Gunathunga, K. K. A. N. N. Perera, L. Ranathunga, P. M. Karunaratne, and T. M. Thanthriwatta, ``Sentiment analysis: Measuring sentiment strength of call centre conversations," in Proc. IEEE ICECCT, Mar. 2015, pp. 1_9.
[9] M. Bouazizi and T. Ohtsuki, "A pattern-based approach for sarcasm detection on twitter," IEEE Access, vol. 4, pp. 5477–5488, September 2016.
[10] R. K. Gupta and Y. Yang, "Crystalnest at semeval-2017 task 4: Using sarcasm detection for enhancing sentiment classification and quantification," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval), Vancouver, Canada, pp. 626– 633, ACL, August 3–4, 2017.
[11]R. G.-I. nez, S. Muresan, and N. Wacholder, "Identifying sarcasm in twitter: A closer look," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL), Portland, Oregon, pp. 581–586, ACL, June 19-24,2011.
[12]B. Pang, L. Lillian, and V. Shivakumar, ``Thumbs up?: Sentiment classification using machine learning techniques," in Proc. ACL-02 Conf.Empirical Methods Natural Lang. Process., vol. 10, pp. 79_86, Jul. 2012.