World Scientific News

An International Scientific Journal

A Comprehensive study: - Sarcasm detection in sentimental analysis

Yamini Ratawal* and **Devendra Tayal**

Indira Gandhi Delhi Technical University for Women, Kashmere Gate, New Delhi - 6, India

*E-mail address: yaminiratawal26@gmail.com

ABSTRACT

Sarcasm detection is one of the active research area in sentimental analysis. However this paper talks about one of the recent issue in sentimental analysis that us sarcasm detection. In our work, we have described different techniques used in sarcasm detection that helps a novice researcher in efficient way. This paper represent different methodologies of carrying out research in this field.

Keywords: Sentimental analysis, Web mining, text mining, opinion mining, machine learning, deep learning

1. INTRODUCTION

Sarcasm is a phenomenon in which the literal meaning of the text or speech is different from the implied sentiments. It is a form of verbal or written irony. The word sarcasm, satire and irony are often used interchangeably but there is a slight difference in their meaning. Irony is when there is a gap between reality and expectations, though irony deployed in order to mock or make fun of someone is sarcasm. Whereas satire is using exaggerated language to make fun of someone. The meaning of sarcasm according to some of the famous dictionaries: - According to Collins English Dictionary, the word sarcasm is speech or writing which actually means the opposite of what it seems to say. It is intended to mock or insult someone.

According to Macmillan Dictionary, sarcasm is the activity of saying or writing to the opposite of what you mean. It is done in order to make someone else feel stupid or show that you are angry. According to Free Dictionary, sarcasm is form of verbal irony intended to express contempt or ridicule.

Sarcasm detection is a fields under sentiment analysis in which people's sentiments, attitude, emotions, feelings are analysed from either text or speech. It is one of the research area that is widely studied in web mining, text mining, data mining and opinion mining. Web mining is divided into three domains: - Web Usage Mining, Web Content Mining and Web Structure Mining. Text mining is done to the data that is extracted and collected from web.

2. TYPES OF SARCASM

Many researchers have classified sarcasm into different categories based on their knowledge and perspective. We have also classified it into broadly five parameters based on Zafaraniet al. [1], Bharti et al. [2], Mondher Bouazizi and Tomoaki Ohtsuki [3].

2. 1. Disparity of sentiments based sarcasm

In this type of sarcasm, the text sentiment conflict with text situation. It is further divided into five parts based on sentiment and situation.

a) Contrast between positive sentiment and negative situation

In this type of sarcasm, there is a contrast between sentimental word and the situation that is depicted in the text. For example, "I love being dead" Or "Awesome, the flight is delayed", these example depict negative situation or activity by using positive sentimental words.

b) Contrast between negative sentiment and positive situation

In this type of sarcasm, sentence include negative word that describe a positive activity resulting into being sarcastic. For example," I dislike Mumbai Indians in IPL because they often win". The negative word in above sentence is "dislike" whereas positive activity is winning the match.

c) Contrasting connotations

In this type of sarcasm, the words used in text have contrasting connotations. For example, "Rohan love being blocked by people on social media"

d) Verity negation

It is a type of sarcasm in which the text contradicts to the fact or universal truth. For example "Sir you are on time today, may be today's sun has risen from west"

e) Temporal facts extraction

This is similar to verity negation but in this text contradicts to fact about an event. For example, "It was nice celebrating my birthday with you and my enemy". Here the event is birthday celebration, contradiction is to celebrate it with your enemy.

2. 2. Conveying emotion as a mean

Sarcasm is used to express emotions based on a person's mood.

a) Wit

In this type of sarcasm, a person will express his unnatural extreme happiness.

In order to express extreme happiness, a person usually uses capital letters, emojis, smileys, question marks etc. Example, "What an AWESOME weather". Here the weather is awesome that is emphasised in this sentence.

b) Snivel

It is used to show a person's angry or annoying mood. It is positive expression in negative situation. For example, "I am pleased that dad woke me up with his annoying loud alarm ringtone". It clearly indicates the angry mood of the person as the word pleased contradict with the situation.

c) Prevarication

It is a type of sarcasm in which a person ignores or avoid to give a clear answer. In this situation, it uses uncommon words, unusual expressions, etc.
Example, "John, you need to come regularly", "Ohhya, I am diamond", here the word diamond is rarely used in this type of situations to give answer to someone.

d) Rampant

It refers to express violence nature of a person by using hyperbolic text.
Example, "Hey Bob, I'm going to need you to work overtime." "Oh, don't worry! I'll be there! Want me to shine your shoes"

2. 3. Sarcasm in written expression

It refers to identify sarcasm based on written expression.

a) Prosodic Variations

In order to emphasize the certain part of text, generally a user repeats letters in a word or punctuations. Commonly seen on social media. For example, "soooooooo", "woooooow", "awesomeeeee", "!!!!!!!!!!!!".

Sometimes in oder to emphasise a text, a user often capitalize the word like "Oh what a AMAZING flower".

b) Structural Variations

In this type of sarcasm, the sarcastic sentence follow a certain pattern like in first part, it will contain opinion of a user and in later part it will describe particular scenario. For example, "I love when my friends don't ask me to go along with them for a party". Here, first part "i love" describes user's view whereas the later part of the sentence describes disappointing situation.

c) Lexical Analysis

This type of sarcasm is purely based on hashtags (#sarcasm, #irony, etc.). For example, "I love being alone" #sarcasm.

2. 4. Sarcasm as function of expertise

In this type of sarcasm, predefined rules of grammar are followed.

a) Language Expertise

Under this category, a user is supposed to be an expert in language that is vocabulary and grammar skills. A user is expected to have a good command over the language. For example, "food is delicious #sarcasm". Here the explicitly added hashtag clearly indicate that food is not good. "food is delicious, it smells like rotten egg". This indicate sarcasm without a hashtag. Person identify sarcasm as he/she has good language skills.

b) Environment Expertise

Users are able to identify sarcasm if they are familiar to their surrounding environment. For example, "I like being called at 3 a.m as my neighbours can't sleep". Here the user is well aware about the surrounding environment.

2. 5. Behaviour-based sarcasm

Behaviour of user is taken into concentration in order to detect sarcastic text.

a) Likes and Dislike Prediction

User may express like or dislike based on various factors like movie reviews, hotel interior or exterior, product, services, events etc.

3. PHASES OF SARCASM DETECTION

The entire process of sarcasm detection follows following steps: **Figure 1.**

3. 1. Data collection

Data collection is one of the inital stage of sarcasm detection. Data could be labeled dataset or unlabeled dataset. If it is unlabeled dataset, then you need to annotate it according to your code. Data can be collected from various sources like twitter blogs, facebook post or ecommerce product reviews etc.

3. 2. Data preprocessing

Next stage is data preprocessing, that involve removal of irrelevant text like hashtags, punctuation marks, slangs, emojis, smileys, etc.
Data preprocessing involve two steps:-

a) Feature extraction

This process involve Pos tagging, term presence, term frequency opinion expression, negation etc.

b) Feature selection

This process include lexicon and statistical method.

3. 3. Apply technique

Next step is to design an algorithm and apply it on the dataset and detect whether the text is sarcastic or not.
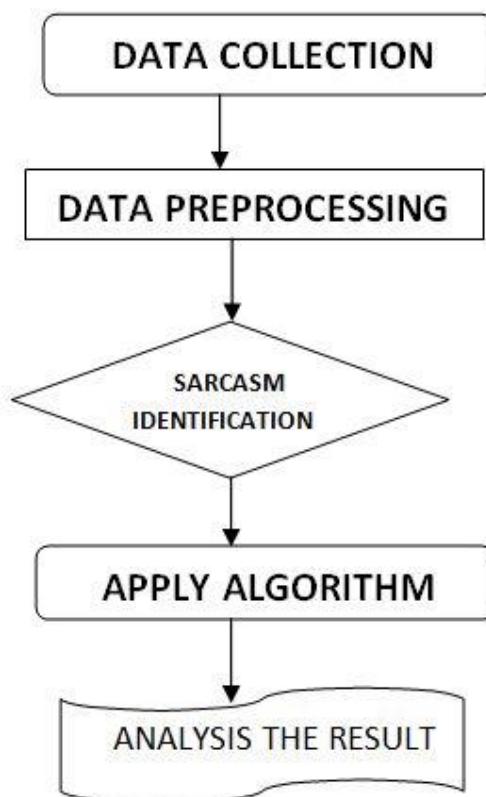

Figure 1. Phases in Sarcasm detection process

4. SARCASM DETECTION TECHNIQUES

There are several approaches in order to detect sarcasm in text.

4. 1. Lexicon based approach

In lexicon based approach, pre-built sentiment dictionaries or lexicons are used. [4] It is based on the assumption that the final orientation of a sentence is equal to the sum of the

individual polarity of words [5]. A parsing based lexical generation algorithm is used to perform sarcasm identification from tweets on Twitter [6].

Mondher Bouazizi at.al [7] proposed an approach to identify the degree of sarcasm found in tweets. Four sets of patterns were used, namely sentiment-related features, punctuation related features, syntactic features and pattern features are extracted using Part of Speech tags. Then Key Performance Indicators (KPIs) is calculated that include, accuracy, precision and recall. It is observed that the precision factor is high for sentiment-related features. Lexicon based approach is used to extract particular features. It is further divided into manual approach, corpus based and dictionary based approach.

a) Manual approach

It is lengthy and time taking process as it is done manually. It is highly labour intensive task.

b) Corpus based approach

Text corpus is a large and structured set of text. It is used to do statistical analysis and hypothesis testing, checking occurrences of words

c) Dictionary based approach

It uses opinion dictionary of positive and negative words that are mapped onto a text in order to find polarity and strength.

4. 2. Machine learning approaches

Setra Genyang [8] has used sarcasm detection with machine learning perspective. Machine learning approaches include three types of learning: - **Figure 2.**
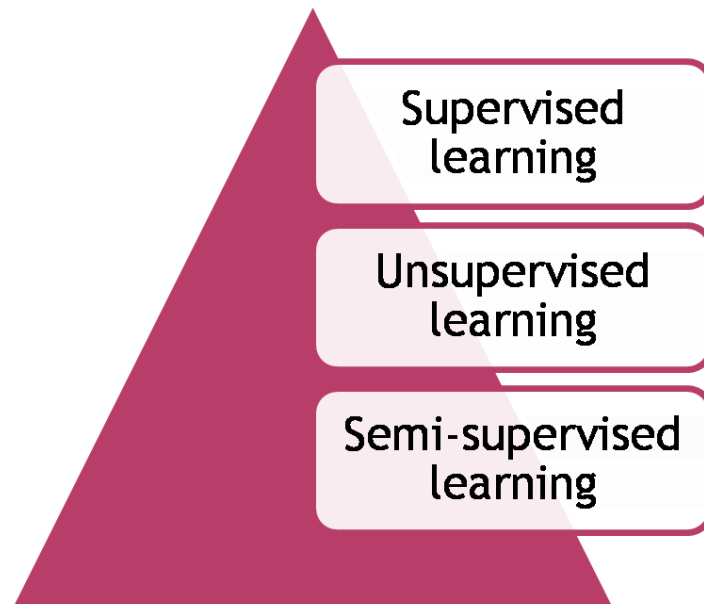

Figure 2. Types of machine learning approaches

a) Supervised learning

A supervised learning algorithm analyses the training or labelled data and produces an inferred function, which can be used for mapping [17].

b) Unsupervised learning

All data is unlabelled and the algorithms learn to inherent structure from the input data.

c) Semi-supervised learning

It uses a combination of labelled and unlabelled dataset. Davidov et al. [9] has proposed a semi supervised approach in his work in order to recognize the sentences as sarcastic or not. The dataset used in his work is Twitter and Amazon dataset.

4. 2. 1. Support vector machine

It is supervised techniques together with learning algorithms that observe data used for classification provided with training examples, which are clearly labelled for belonging to one of the types. An SVM training algorithm develops a system that assigns unique examples to each group, making it a non-probabilistic, binary linear classifier [10].

4. 2. 2. Naive Bayes classifier

It is based on Bayes theorem. It assumes that the value of a specific feature is independent of the value of any other feature. This classifier is straight forward, uncomplicated and efficient for large datasets, without any complex iterative parameter estimation. Brandon Joyce [11] has used naive Bayes classifier in order to analysis Tweets for the 2016 US Presidential Election.

4. 2. 3. Maximum entropy

It is a classifier which is based on probability distributions of the data. The primary rule is that when no information is known then the distribution should have maximal entropy [12]. The labelled training data offers restriction on the distribution and find out where to have minimal non-uniformity. Soni et al. [13] devised twitter data by using classification algorithm like maximum entropy and got best accuracy by using combination of max entropy and google translator.

5. CHALLENGES IN SARCASM DETECTION

Opinion faking is one of the challenge as certain organizations and companies take benefit of the fact that the user's identity is anonymous on the web. They place spam reviews to discredit other's products. Variations in spelling is another challenge as casual setting and limitations on text length, syntax rules, and spellings are seldom taken care of by the people. Multilingual condition is one of the challenge for researchers as the people should have different language skills for this purpose.

6. FUTURE SCOPE AND CONCLUSION

There is a vast future scope in this research area as it is not only interesting but also challenging in real world. Perplexed Bayes classifier can be used as Naive Bayes classifiers estimate posterior probabilities poorly. Naive Bayes assumes conditional independence whereas perplexed Bayes classifier is dependent. Further the work can be done on incongruity in numbers [18] classifier can be used as Naive Bayes classifiers estimate posterior probabilities poorly. Naive Bayes assumes conditional independence whereas perplexed Bayes classifier is dependent. Further the work can be done on incongruity in numbers [18-20]

References

[1]   Ashwin Rajadesingan, Reza Zafarani, and Huan Liu, Sarcasm detection on twitter: A behavioural modeling approach. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, ACM, 2015.

[2]   Santosh Kumar Bharti, Korra Sathya Babu, and Sanjay Kumar Jena, "Parsing-based sarcasm sentiment recognition in Twitter data. 2015IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2015.

[3]   Mondher Bouazizi and Tomoaki Otsuki. A pattern-based approach for sarcasm detection on twitter. *IEEE Transl.* 2016.

[4]   T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space. Proc. Int. Conf. Learning Representations, 2013. arXiv:1301.3781

[5]   F. Morin and Y. Bengio, "Hierarchical probabilistic neural network language model. Proc. Int. Workshop Artif. Intell. Statist. 2005, pp. 246–252.

[6]   S. K. Bharti, K. S. Babuve S. K. Jena, Parsing-based Sarcasm Sentiment Recognition in Twitter Data. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Paris, 2015.

[7]   Mondher Bouazizi, Tomoaki Ohtsuki, Sarcasm Detection in Twitter. Global Communications Conference (GLOBECOM), 2015 IEEE.

[8]   Setra Genyang Wicana, Taha Yasin İbisoglu, Uraz Yavanoglu, A Review on Sarcasm Detection from Machine-Learning Perspective. 2017 IEEE 11th International Conference on Semantic Computing

[9]   D.Davidov, Dmitry, O. Tsur, A. Rappoport. Semisupervised Recognition of Sarcastic sentences in Twitter and Amazon. Proc. 14th Conf. on Computational Natural Language Learning, pp. 107116, 2010

[10]   Shweta Rana, Archana Singh, Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. 2nd International Conference on Next Generation Computing Technologies (NGCT) 2016

[11] B. Joyce and J. Deng, Sentiment analysis of tweets for the 2016 US presidential election. 2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*,* 2017, pp. 1-4, doi: 10.1109/URTC.2017.8284176

[12] Bamman, David, and Noah A. Smith. Contextualized Sarcasm Detection on Twitter. Ninth International AAAI Conference on Web and Social Media. 2015

[13] A. K. Soni, Multi-lingual sentiment analysis of Twitter data by using classification algorithms. 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*,* 2017, pp. 1-5, doi: 10.1109/ICECCT.2017.8117884

[14] Aditya Joshi, Vaibhav Tripathi, Pushpak Bhattacharyya, Mark Carman, "Harnessing Sequence Labeling for Sarcasm Detection in Dialogue from TV Series 'Friends' "Conference on Computational Natural Language Learning (CoNLL), pages 146–155, Berlin, Germany, August 7-12, 2016.

[15] K. S. Krishnaveni, R. R. Pai and V. Iyer, Faculty rating system based on student feedbacks using sentimental analysis. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*,* 2017, pp. 1648-1653, doi: 10.1109/ICACCI.2017.8126079

[16] Lohita Sivaprakasam, Aarthe Jayaprakash, A Study on Sarcasm Detection Algorithms, *International Journal of Engineering Technology Science and Research* Volume 4, Issue 9 September 2017

[17] Han-Xiaoshi, Xiao-Junli, A sentiment analysis model for hotel reviews based on supervised learning, International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011

[18] Aditya Joshi, Vinita Sharma, Pushpak Bhattacharyya, Harnessing context incongruity for sarcasm detection, 7th International Joint Conference on Natural Language Processing, Vol. 2. 757-762

[19] Magistry, P., Hsieh, SK. & Chang, YY. Sentiment detection in micro-blogs using unsupervised chunk extraction. *Lingua Sinica* 2, 1 (2016). https://doi.org/10.1186/s40655-015-0010-8

[20] M. Korakakis, E. Spyrou and P. Mylonas, A survey on political event analysis in Twitter. 2017 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*,* 2017, pp. 14-19, doi: 10.1109/SMAP.2017.8022660

[21] Duyu Tang, Furu Wei, Bing Qin, Nan Yang, Ting Liu, and Ming Zhou, Sentiment Embeddings with Applications to Sentiment Analysis, IEEE Transactions on Knowledge and Data Engineering, Vol. 28, no. 2, February 2016