Inspection of Retrospection - Challenges in Sentiment Analysis

Ms. Puja M. Dadhe[1] and Dr. R.N. Jugele[2]
[1] Research Scholar, Department of Computer Science,
Shivaji Science College, Nagpur.
[2]Assosiate Professor, Department of Computer Science,
Shivaji Science College, Nagpur.
poojadadhe@gmail.com[1], rn_jugele@yahoo.com[2]

Abstract: *With the advancement in social media and technology, social media had became a platform for millions of users' to share views and opinion related to products, issues and policies. Data generated through this media is an important source for analysis and opinion mining for extracting sentiments. Performing such analysis is not an easy task, it encounters multiple challenges which need to be addressed. This paper presents a retrospection of various challenges of sentiment analysis.*

Keywords: **Sentiment analysis, Social media, opinion mining, Sentiment polarity, NLP, summarization, Negative, Positive.**

1. INTRODUCTION

Sentiment Analysis [12] is a branch of Natural language Programming intended to mine various sources of data for opinions. It is defined as a computational study of human's thoughts or opinions, emotions and attitudes toward an object [2]. It is done on the data that is collected from the Internet and various social media platforms. Organizations, Companies, and Governments often use sentiment analysis to understand how the people feel about themselves, products and their policies. The purpose of Sentiment analysis is to classify the polarity of user's sentiment and extract his opinion regarding an interested entity, which help in providing valuable information for decision making [3]. Polarity in sentiment analysis means classifying the sentiments as positive, negative and neutral. Sentiment analysis has been classified into different levels:

- **Document level**: This classifies the whole document text into positive or negative polarity.
- **Sentence level**: Which extracts the polarity of each sentence of a document into positive or negative polarity.
- **Aspect/entity level**: Which classify the sentiment polarity of each entity's aspect or feature of a sentence/document [4].

There are two kinds of information in a particular sentence:
Objective- An objective sentence states factual information about the world.
Subjective- A subjective statement expresses some personal feeling, belief or view.
The task of determining whether a sentence is subjective or objective is called subjectivity classification.

The resulting subjective sentences are further classified as:
Expressing positive or negative opinions. This is called as sentence level sentiment classification [11].

Sentiment Analysis involves various aspects as shown in Figure 1. First phase involves data collection, various data from different sources like blogs, reviews and microbloging (Twitter, Facebook) which act as an input for Sentiment Analysis. Pre-Processing phase involves cleaning and Scraping of data. Feature Extraction identifies aspect which are being referred in particular sentence, document or comment by customer. Sentiment Analysis approaches are the techniques used for classification of Sentiment. Last phase is to generate results.

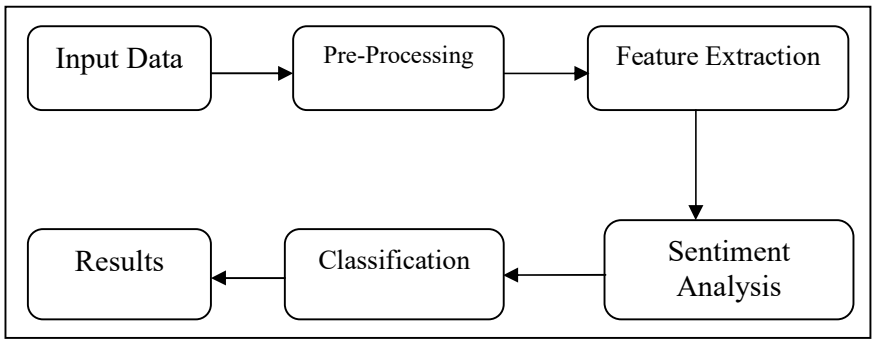

Figure 1- An Aspect of Sentiment Analysis.

Standard Structure of Sentimental Analysis:
Sentiment Analysis concludes whether users' view is positive, minus/neutral about a product, issue, event, etc. It is describe in three primary steps [1].

- **Data Retrieval** - It is the procedure of collecting review text from review sites. Different review websites contain reviews for products, movies, hotels and news. Also include Information retrieval -Techniques such as web crawler can be employed to collect the review text data from many sources and store them in a database. This step involves retrieval of reviews, micro-blogs and comments by user [1].

- **Sentiment Classification** - Primary steps in sentiment analysis are a classification of review text. Given a review document M = {M1….. M1} and a predefined category set K = {positive, negative}, sentiment classification is to classify each day in M, with a label expressed in K. The approach involves classifying review text into two forms namely positive and negative. Machine learning and dictionary based approach is more popular [1].

- **Sentiment Summarization** - Summarization of Sentiment is a major character in the Sentiment Analysis process. Summary of reviews should be based on features or subtopics that are mentioned in the reviews. Many works have been done on summarization of product reviews [1].

2. Techniques of Sentiment Analysis

The techniques of Sentiment classification [8] is divided into:

- Lexicon based approach.
- Machine learning approach.
- Hybrid approach.

2.1 Supervised Machine Learning:

Classification is most frequently used popular data mining technique. It is used to predict the possible outcome from given data set on the basis of defined set of attributes and a given predictive attributes. The given dataset is found to be the training dataset consist on independent variables (dataset related properties) and a dependent attribute (predicted attribute). A training dataset created model test on test corpus contains the same attributes but no predicted attribute. Accuracy of model checks that how accurate it is to make prediction. Product features and sentenced words are extracted using Double Propagation Algorithm [8].

2.2 Unsupervised Learning

In contrast of supervised learning, unsupervised learning has no explicit targeted output associated with input. Class label for any instance is unknown so unsupervised learning is about to learn by observation. Clustering is technique used in unsupervised learning. The process of gathering objects of similar characteristics into a group is called clustering. Objects in one cluster are dissimilar to the objects in other clusters [8].

2.3 Case Based Reasoning

Case based reasoning is an emerging Artificial Intelligence supervised technique. It is a powerful tool of computer reasoning and solve the problems (cases) which is closest to real time scenario. It is a problem solving technique in which knowledge is personified as past cases in library and it does not depend on classical rules. The solutions are stored in CBR repository called Knowledge base or Case base [8].

3. Literature Review

In 2015, P.Kalarani, Dr.S. Selva Brunda, examined various Sentiment Analysis challenges. They discussed about the Challenges and application area in opinion mining and the techniques and tools used for opinion mining. Challenges discussed by them included detection of spam, fake reviews, limitation of classification filtering, asymmetry in availability of opinion mining software, incorporation of opinion with implicit and behaviour data, domain-independence and natural language processing overheads[8].

In 2016, Mohey El-Din, Doaa, discussed the importance and effects of sentiment analysis challenges in sentiment evaluation based on two comparisons among forty-seven papers they reviewed. They recognised the sentiment challenge of domain-dependence. They also suggested that the negation challenge was the popular in all types of review structure. They found the relationship between the

proportion of sentiment techniques usage in theoretical and technical types to solve sentiment challenges [6].

In 2017, Osamah A.M Ghaleb,Anna Saro Vijendran, out looked concept of Sentiment Analysis and presented multiple challenges which includes
1. Big Data-related Issues like data collection, data pre-processing and storage.
2. Language-oriented Issues - lack of lexicon, different writing style and different word meaning.
3 Fake opinions- It includes Fake Positive or Fake Negative opinions.
4. Text related issues [7].

In 2018, Syed Saood Zia , Sana Fatima , IdrisMala , M. Sadiq Ali Khan , M. Naseem , Bhagwan Das, published the biggest challenge faced in sentiment analysis is the domain specific nature of opinionated words. They may perform well in one domain but works poor in other domain. Since every human being has a different nature so it is very hard to correctly classify users provided input belongs to a particular entity [10].

In 2019 S. V. Pandey and A. V. Deorankar, covered different levels of sentiment analysis and discussed aspect-based sentiment analysis. They proposed important challenges to this research area like named entity recognition, sentiment polarity detection, subjectivity detection etc. described with suitable example. They used Stanford Core NPL tools to visualize the result of some basic operation of NLP which can be used for sentiment analysis [9].

4. Sentiment Analysis Challenges

Sentiment Analysis seemed to be just classification problem but more we dive in it more challenging it appears. Challenges can be categorized as linguistic challenges and accuracy based challenges.
4.1 **Linguistic Challenges**- These challenges are related to the language (basically English) used by the author [13]
Feature recognition- This challenge is more related to lack of context in the piece of data being analyzed. For example the statement “it was beautiful” lacks the subject in discussion.
Sarcasm- Sarcastic statements are those in which people express negative thoughts using positive words. This situation occurs usually on social media where one put up opinion in much sarcastic ways which can easily confuse sentiment analysis model.
Contradiction- people can be contradictory in the way they review any product or give their opinion.
Slangs- People may use words like “OMG” ,“TC” or “LOL” to express their response. Identifying these words needs extra efforts to train our network to correctly identify the sentiment.
Jokes- Human brain can easily understand jokes but it is harder for a computer to parse.
Irony- Similar to sarcasm irony statement also causes misinterpretation of data and poses problems to natural language processing.

Emoticons- Massive use of emoticons (emotion icon, emoji) that is pictorial representation of human expression using graphical icon used to express mood or feelings are hard to interpret.

4.2 **Accuracy Based** [6] – This involves challenges which affects to the accuracy of models used to classify sentiments. These are sub-divided into two types of challenges theoretical and technical challenges.

Huge lexicon- This is theoretical challenge, require large amount of lexicon resources. With the advent new words every day it has became hard to handle huge lexicon.

Bi-polar- This is technical challenge, words can be positive and negative at the same time with regards to the context been used.

NLP overhead(emotions)- This is technical challenge, The natural language overhead like ambiguity, co-reference, Implicitness, inference etc. created hindrance in sentiment analysis too[8].

Domain Dependence- This is theoretical challenge, which requires networks to be trained on the specific domain they are analyzed for.

Spam and Fake- This is theoretical challenge, duplicate and similar reviews are fake reviews. This results in wrong and invalid analysis of data for decision making.

Negation- This is theoretical challenge, Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very different from "the picture was not great"[5].

5. Conclusion

Sentiment Analysis field has been proven to be most promising research field in data analytics. Providing ranges of solutions for decision making and become popular platform in mining domain. Sentiment Analysis job involves number of challenges which need to be properly addressed for to perform accurately. This paper retrospect's the concept and challenges related to Sentiment Analysis. More Research work in foreign languages other than English is essential. Many challenges related to slang, inferences, ambiguity and contradiction needs more précised techniques to be overcome properly. These challenges provide new areas for further research.

REFERENCES

Journal Article

*[1] Abhishekh Kaushik,Anchal Kaushik,Sudhanshu Naithani, "A Study on Sentiment Analysis: Methods and Tools", International Journal of Science and Research, Volume 4 Issue 12,**2015.***

*[2] Chandni, Navchandra, "Sentiment Analysis and its Challenges", International Journal of Engineering Research and Technology, Vol 4,issue 0, March **2015**.*

*[3] Haseena Rahmath P, "Opinion Mining and Sentiment Analysis -Challenges and Applications", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 5, May **2014**.*

*[4] Doaa mohey el-din Mohamed Hussein,"A survey on sentiment analysis challenges", Journal of King Saud University - Engineering Sciences, volume 30,issue 4, October **2018**, pages 330-338.*

*[5] Dr. P. Sumathy, S. M. Muthukumari."Sentiment Analysis of Twitter Data Using Multi Class Semantic Approach",International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Volume 3,Issue 6, **2018**.*

[6] Mohey El-Din, Doaa." A Survey on Sentiment Analysis Challenges", Journal of King Saud University - Engineering Science. 10.1016/j.jksues.2016.04.002.

*[7] Osamah A.M Ghaleb ,Anna Saro Vijendran, "The Challenges Of Sentiment Analysis On Social Web Communities", International Journal of Advance Research in Science and Engineering,Voulme No.06,Issue No. 12, **2017**.*

*[8] P.Kalarani, Dr.S. Selva Brunda," An Overview on Research Challenges in Opinion Mining and Sentiment Analysis", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 10, October **2015**.*

*[9] S. V. Pandey and A. V. Deorankar, "A Study of Sentiment Analysis Task and It's Challenges", 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, **2019**.*

*[10] Sana Fatima, Syed Saood Zia , IdrisMala , M. Sadiq Ali Khan , M. Naseem , Bhagwan Das,"A Survey on Sentiment Analysis, Classification and Applications", International Journal of Pure and Applied Mathematics Volume 119 No. 10 **2018**.*

*[11] Sweta Morajkar , Maria Christina Barretto1, "Sentiment Analysis Tools and Techniques: A Comprehensive Survey", International Journal for Research in Applied Science & Engineering Technology (IJRASET),Volume 5 Issue XI November **2017**.*

Links

[12] https://en.wikipedia.org/wiki/Sentiment_analysis

[13] https://www.toptal.com/deep-learning/4-sentiment-analysis-accuracy-traps