SENTIMENT ANALYSIS CHALLENGES: A REVIEW

Chandrashekhar Tople and P. B. Dhumne

Sardar Patel Mahavidyalaya Chandrapur
Corresponding Email: cstopale@gmail.com, pdhumane@rediffmail.com

ABSTRACT:

Reviews, opinions, suggestions, ratings, and feedback are produced by writers as websites, social networks, blogs, and online portals potentially generate. For organizations, governments, and people in general, these attitudes are extremely advantageous. Even if the majority of this writer-generated information has to be mined using text mining algorithms and sentiment analysis, it is still considered to be informative. But the process of sentiment analysis and assessment faces a number of difficulties. These difficulties create roadblocks in determining the correct sentiment polarity and accurately evaluating the meaning of feelings. Sentiment analysis is the process of identifying and extracting subjective information from text using text analysis and natural language processing tools. In this context, this paper gives a thorough assessment of the relevant literature with a specific focus on challenges associated in various application domains for sentiment analysis.

Keywords: - Sentiment Analysis, NLP, Machine Learning.

INTRODUCTION :

Due to the fast digitalization over the past ten years, individuals now utilize the internet to complete all of the fundamental duties that they formerly completed manually. It is more challenging to determine or comprehend the product quality as e-commerce websites use photographs to advertise the items.

As a result, a facility for rating and writing reviews of items exists. Simply said, consumer ratings of items just reflect how likely they are to be liked overall. The other feature, reviews, aids the potential customer in fully comprehending the product. There are many reviews available for each product because there are millions of users who access different e-commerce websites. Even if it were possible, it would be very challenging for a human to analyze all of the reviews, making it impossible for potential customers to read them all. Before making a purchase, people frequently consult the reviews left by friends or family who have already purchased the item. Reviews and ratings of things are very important in today's society for forming opinions.

To solve these issues It employs sentiment analysis. Sentiment analysis makes it simple to analyse the emotions included in a statement. Natural Language Processing(NLP) may be used to automate laborious operations like reviewing reviews. It may be used for a variety of things, like identifying spam emails, unethical tweets, phone reviews, identifying product characteristics from reviews, etc. Information may be separated into subjective and objective categories. While subjective information focuses on feelings, objective information explains facts. Sentiment analysis makes use of subjective data.

There are two primary methods for sentiment analysis: machine learning and lexicon-based methods and hybrid methods. More information on sentiment analysis classification is provided in Fig. 1.

By manually preparing, the sentiment analysis [3] problem can be resolved to an alluring degree. However, there hasn't yet been a system for sentiment analysis that is entirely automated and doesn't require manual mediation. Fundamentally, this is due to the challenges in this industry. This essay attempts to provide a writing analysis of the subject of slant research and supposition mining. This study offers a glimpse into a few of the key reviews that have emerged in the field.

METHODOLOGY :

As previously indicated, sentiment analysis employs both lexicon-based analysis and machine learning. Data gathering, data preparation, lexicon-based analysis, and processed output are only a few of the procedures that are involved.

Sentiment Classification Techniques :

The three approaches to sentiment analysis are lexicon-based, machine learning-based, and hybrid approaches[2]. The entire manuscript is divided into lexemes using the lexicon-based technique, which is then utilised to analyse the sentences. Corpus-based approach and dictionary-based approach are further categories for lexicon-based approaches.

The polarity of the statement is identified as negative, positive, or neutral using a corpus-based technique. Negative: awful, repulsive, bothersome, etc. and Positive: best, lovely, wonderful, etc.

A dictionary-based technique is a quantitative way to gauge how the reader is feeling about the sentences.

Unsupervised learning and supervised learning are two types of machine learning approaches that use machine learning algorithms. The desired and actual outcomes must be compared during supervised learning. Unsupervised learning uses data and prior experience to increase its accuracy rather than requiring any

desired outcome. Hybrid approaches combine the approaches of machine learning and Lexicon-based approaches.

CHALLENGES :

Sarcasm detection :

Sarcasm is described by the Macmillan English dictionary as the act of saying or writing the reverse of what one intends, or of speaking in a way meant to make another person feel foolish or demonstrate his anger to them [6,7]. When someone writes something good but really means negative or vice versa, the problem of sarcasm in sentiment analysis arises, which makes the process of sentiment analysis more difficult.

In everyday conversation, we frequently utilise sarcastic terms. As a result, there is a growing interest in sarcasm detection to solve the issue of obtaining dishonest feelings by automatically recognising sarcastic statements in a given text. Sarcasm identification is an extremely difficult NLP job because of the intricacy and ambiguity of sarcasm [8].

Sarcasm detection methods come in a variety of forms [8,9]. Deep learning was utilised by Jain et al. [10] to identify sarcasm in real-time when English and an Indian local language were combined (Hinglish).

Negation handling :

Negation words, such as not, neither, or, etc., must be handled carefully since they might change the polarity of a paragraph. For instance, the line "This movie is good" would be considered a positive sentence whereas the sentence "The movie is not good" would be considered a negative sentence.

Sadly, some systems exclude negation words because they are on Stop-Word lists or because they are implicitly ignored since they have a neutral emotion value in a lexicon and do not affect the final polarity. Negation words can be discovered in sentences without changing the

tone of the text, hence it is difficult to perform this assignment by flipping the polarity.

Lazib et al. [11] suggested a hybrid neural network based on syntactic paths for the identification of negation scope.

Spam detection :

Spam identification is crucial in the realm of sentiment analysis.. Spam and phoney reviews can harm a brand's reputation and artificially sway consumers' impressions of goods, services, businesses, or other organisations since online opinions affect consumer buying choices [12]. It is quite difficult to create a spam detection system that can spot phoney reviews amid a large number of evaluations since there isn't a clear distinction between them. A system created by Saumya and Singh [13] is one of the methods that was suggested to carry out the work of spam identification. This system effectively uses three features: the sentiment of the review and its comments, the content based factor, and rating deviation.

Anaphora and coreference resolution :

A link of coreference between linguistic concepts is an anaphora [14]. It is helpful to know what a pronoun in a phrase refers to in sentiment analysis, especially for aspect-based, since it aids in extracting all the facets of a given item.

Pronouns are unfortunately frequently disregarded or dropped during the preprocessing stage. A thorough survey of the fields of coreference resolution and the closely related field of anaphora resolution was provided by Sukthanker et al. [15].

Word sense disambiguation (WSD) :

A word may signify multiple things depending on the occasion and the context as well as the area in which it is employed. The goal of word sense disambiguation is to identify the sense in which a word has been used in a phrase. For instance, the term "curved" denotes a good context when used in relation to television yet may denote a negative meaning when used in relation to a mobile phone. Therefore, it might be quite difficult to extract a word sense from a phrase. Wang et al. [17] suggested a knowledge-based approach that uses the well-known lexicon WordNet to resolve this difficult job.

Low-resource languages :

The majority of research studies in the field of sentiment analysis have concentrated on the English language [4] or other languages that have a sufficient number of linguistic resources (such as a sentiment lexicon and tagged text corpus). As was previously noted, supervised learning techniques are most frequently employed for sentiment analysis. These methods, however, significantly rely on linguistic resources, which can be expensive to acquire for uncommon languages [5]. Low-resource languages (or under resourced languages) are the categories of languages that experience linguistic resource shortage.

Sentiment analysis of code-mixed data :

Use of vocabulary and grammar from several languages in a single phrase is known as code-mixing (CM) [19,20]. It is extremely typical in communities that speak many languages and presents a significant obstacle for NLP tasks like sentiment analysis. The identification of compositional semantics, which are crucial for conducting sentiment analysis utilising rule- and machine learning-based approaches, is hampered by the lack of a formal grammar for code-mixed phrases.

In addition, since mixing is up to the individual, there are no set guidelines for mixing, which is one of the biggest challenges [20].

CONCLUSION :

Sentiment analysis is a machine learning problem that has recently generated a lot of interest in academic research. The relevant writings done to understand sentiment analysis challenges may be taken into consideration through this writing review. Although there have been a few notable efforts in this area, a fully

automated and incredibly effective framework has not yet been offered.

In the area of machine learning, there are a few issues that still need to be resolved, including element acknowledgment, co-reference resolution, space dependency, and others. These problems were dealt with separately, and those arrangements might be used to improve assumptions investigation processes.

REFERENCES:

Anees, A. F., Shaikh, A., Shaikh, A., & Shaikh, S. (2020). Survey paper on sentiment analysis: Techniques and challenges. EasyChair2516-2314.

Diana Maynard, Adam Funk. Automatic detection ofpolitical opinions in tweets. In: Proceedings of the 8thinternational conference on the semantic web,ESWC'11; 2011.

Raisa Varghese and Jayasree M" A survey on sentiment analysis and opinion mining" IJRET Vol.02, Issue no.11, pp.312-317, Nov 2013

M.S. Hajmohammadi, R. Ibrahim, A. Selamat, H. Fujita, Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples, Inf. Sci. (Ny). 317 (2015) 67–77, https://doi.org/10.1016/j.ins.2015.04. 003.

Y. Ren, N. Kaji, N. Yoshinaga, M. Kitsuregawa, Sentiment classification in under-resourced languages using graph-based semi-supervised learning methods, IEICE Trans. Inf. Syst. E97.D (2014) 790–797, https://doi.org/10. 1587/transinf.E97.D.790.

M. Rundell, Macmillan English Dictionary for Advanced Learners, 2007, pp. 1–1748, https://books.google.co.ma/books?id=v 6SISQAACAAJ.

M. Birjali, A. Beni-Hssane, M. Erritali, A method proposed for estimating depressed feeling tendencies of social media users utilizing their data, in: Adv. Intell. Syst. Comput., 2017, pp. 1–8, https://doi.org/10.1007/978-3- 319-52941-7_41.

L. Ren, B. Xu, H. Lin, X. Liu, L. Yang, Sarcasm detection with sentiment semantics enhanced multi-level memory network, Neurocomputing. (2020) 1–7, https://doi.org/10.1016/j.neucom.2020 .03.081.

Y. Ren, D. Ji, H. Ren, Context-augmented convolutional neural networks for twitter sarcasm detection, Neurocomputing. 308 (2018) 1–7, https: //doi.org/10.1016/j.neucom.2018.03.0 47.

D. Jain, A. Kumar, G. Garg, Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN, Appl. Soft Comput. 91 (2020) 1–11, https://doi.org/10.1016/j.asoc.2020.10 6198.

L. Lazib, B. Qin, Y. Zhao, W. Zhang, T. Liu, A syntactic path-based hybrid neural network for negation scope detection, Front. Comput. Sci. 14 (2020) 84–94, https://doi.org/10.1007/s11704-018- 7368-6.

E.F. Cardoso, R.M. Silva, T.A. Almeida, Towards automatic filtering of fake reviews, Neurocomputing 309 (2018) 106–116, https://doi.org/10.1016/j. neucom.2018.04.074.

S. Saumya, J.P. Singh, Detection of spam reviews: a sentiment analysis approach, CSI Trans. ICT. 6 (2018) 137–148, https://doi.org/10.1007/ s40012-018-0193-0.

I. Toledo-Gómez, E. Valtierra-Romero, A. Guzmán-Arenas, A. CuevasRasgado, L. Méndez-Segundo, AnaPro, tool for identification and resolution of direct anaphora in Spanish, J. Appl. Res. Technol. 12 (2014) 14–40, https://doi.org/10.1016/S1665-6423(14)71602-5.

R. Sukthanker, S. Poria, E. Cambria, R. Thirunavukarasu, Anaphora and coreference resolution: A review, Inf. Fusion. 59 (2020) 139–162, https://doi.org/10.1016/j.inffus.2020.01.010.

L.J. Deborah, V. Karthika, R. Baskaran, A. Kannan, Enhanced Anaphora Resolution Algorithm Facilitating Ontology Construction, 2011, pp. 526–535, https://doi.org/10.1007/978-3-642-22555-0_54. [17] Y. Wang, M. Wang, H. Fujita, Word sense disambiguation: A comprehensive knowledge exploitation framework, Knowledge-Based Syst. 190 (2020) 1–13, https://doi.org/10.1016/j.knosys.2019.105030.

Y. Xia, E. Cambria, A. Hussain, H. Zhao, Word polarity disambiguation using Bayesian model and opinion-level features, Cognit. Comput. 7 (2015) 369–380, https://doi.org/10.1007/s12559-014-9298-4. [19] C. Myers-Scotton, Common and uncommon ground: Social and structural factors in codeswitching, Lang. Soc. 22 (1993) 475–503, https://doi.org/10.1017/S0047404500017449.

S. Poria, D. Hazarika, N. Majumder, R. Mihalcea, Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research, IEEE Trans. Affect. Comput. (2020) 1, s.
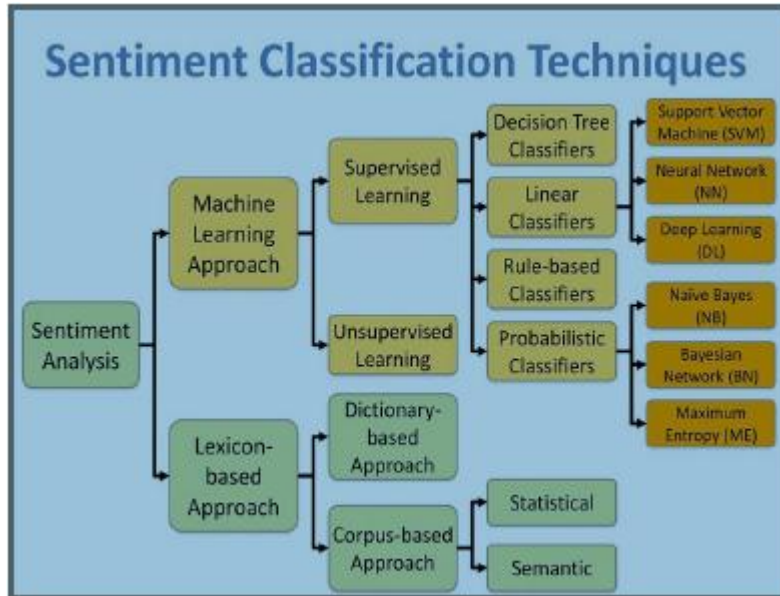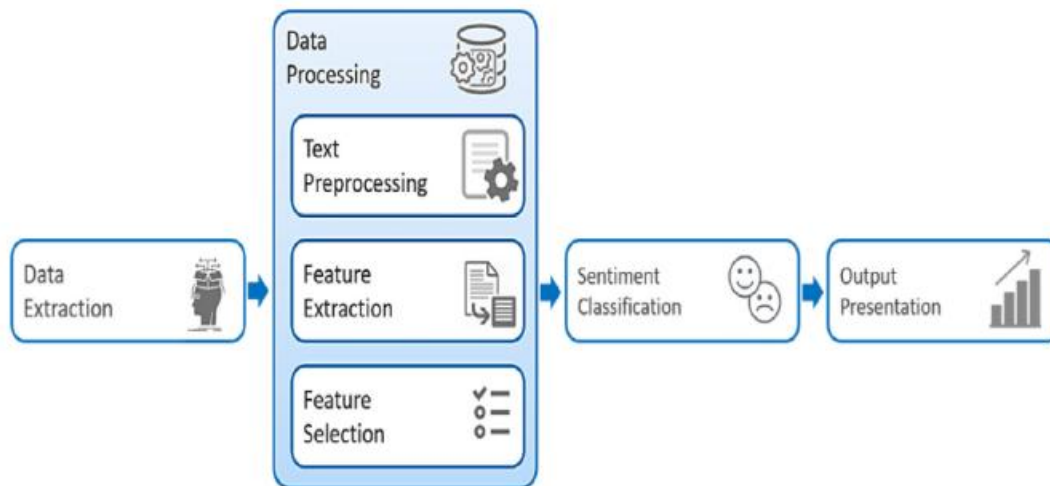
Fig. 1[1]

Fig 2. The generic process of sentiment analysis.