MODELLING OF CONTEXT-BASED SENTIMENT ANALYSIS AND ITS APPLICATIONS

A THESIS

SUBMITTED TO THE DELHI TECHNOLOGICAL UNIVERSITY

FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMPUTER ENGINEERING

SUBMITTED BY

GEETANJALI GARG

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

DELHI- 110042 (INDIA)

2020

MODELLING OF
CONTEXT-BASED SENTIMENT ANALYSIS
AND
ITS APPLICATIONS

BY

GEETANJALI GARG

SUBMITTED TO THE DELHI TECHNOLOGICAL UNIVERSITY IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMPUTER ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

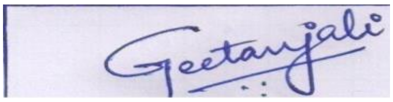
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
DELHI- 110042 (INDIA)

2020

CANDIDATE DECLARATION

I hereby certify that the research work which is being presented in this thesis entitled **"Modelling of Context-Based Sentiment Analysis and Its Applications"** in fulfilment of requirements of the award of degree of Doctor of Philosophy, is an authentic record of my own research work carried out under the supervision of Dr. Akshi Kumar.

The matter presented in this thesis has not been submitted elsewhere in part or fully to any other University or Institute for award of any degree.



Geetanjali Garg

2K16/PHD/CO/05
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India-110042.

DELHI TECHNOLOGICAL UNIVERSITY

(Govt. of National Capital Territory of Delhi)

BAWANA ROAD, DELHI – 110042

CERTIFICATE

Date: 9-11-2020

This is to certify that the thesis entitled "**Modelling of Context-Based Sentiment Analysis and Its Applications "** done by Geetanjali Garg, Roll no. 2K16/PHD/CO/05 in Department of Computer Science & Engineering, Delhi Technological University is an authentic work carried out by her under my guidance.

This work is based on original research and the matter embodied in this report has not been submitted earlier for the award of any degree or diploma to the best of my knowledge and belief.

Supervisor

Akshi Kumar

Dr. Akshi Kumar
Assistant Professor
Department of CSE, DTU

ACKNOWLEDGEMENT

I feel pride in placing on record my deep gratitude to my Ph.D. supervisor, Dr. Akshi Kumar, Assistant Professor, Department of Computer Science and Engineering, Delhi Technological University, Delhi, who guided me throughout this long journey of four years of research work. Whether it was review of literature or clearing doubts, she spared quality time to guide me despite her busy schedule. She not only helped in solving all problem with comfortable ease but also encouraged and motivated me to sail through in difficult times. I have gained immensely from her comments and suggestions regarding the technical quality of the work. It is hard to imagine successful completion of such a research work without her guidance and care.

I am very grateful to Prof. Rajni Jindal, Head, Department of Computer Science & Engineering, Delhi Technological University, Delhi, for her constant encouragement and support to accomplish this task.

I am thankful to all the other faculty members of Department of Computer Science and Engineering, Delhi Technological University, Delhi, for the motivation and inspiration. I am also thankful to all non-teaching staff of CSE Department, DTU who have helped me directly or indirectly in completion of this research work.

I wish to pay high regards to my parent in-laws for their invaluable support, best wishes and encouragements. I am grateful to my husband and children for their love, patience and support for completing this project. Without them this research would not have been possible.

Words are not enough to express my gratitude to my dear father, mother and brother for their unconditional love and innumerable sacrifices. I thank them for making me capable enough to undertake and finish this challenging task and for providing everything I ever asked for.

Geetanjali Garg
2K16/PHD/CO/05 (PhD Scholar)
Department of Computer Science & Engineering
E-mail: geeta.anjali.garg@gmail.com

ABSTRACT

"You shall know a word by the company it keeps"

----- *Firth, J. R. 1957*

Social media mining has proven valuable in numerous research areas as a pragmatic tool for public opinion extraction and analysis. Sentiment analysis (SA) addresses the dynamics of complex socio-affective applications that permeate intelligence and decision making in the sentient and solution-savvy social web. It encompasses investigation of both opinion and emotion within the content. Having started with simple polarity detection, contemporary SA has advanced to a more nuanced analysis of context, affect and emotion sensing. Existing sentiment analysis techniques quite efficiently capture opinions from text written in syntactically correct and explicit language. However, while dealing with the informal social data, limitations have been observed in performance of sentiment analysis techniques. Understanding the pragmatics, emotion, cognition and behaviour are key to accurate SA. Ongoing research shows that some of the issues pertaining to natural language use can be resolved by adding extra information (i.e. context) to the process of SA.

This research primarily aims to find out the types of contextual information which can be extracted from social media and can be applied to improve results of SA. In this direction, a multi-faceted conceptual framework for context has been built. It defines types of context that can be used in SA. The concept of context was then applied for building a model for contextual SA. In addition to this work, the contextual framework work dealing with single modality (textual data) has been extended to deal with multiple modality data. This conceptualization of 'context' was further applied for detection of specialized sentiment like sarcasm with improved accuracy. Sarcasm detection has been carried out for both multiple modality data and multilingual data. This work presents a learning model for real-time sarcasm detection in Hinglish (Hindi +English) code-switch dataset. The empirical analysis has been carried on the available benchmark datasets and also on created datasets. The results have been evaluated using standard classification metrics and proposed techniques have been compared extensively with the existing state-of-the-art. The research affirms that the use of appropriate context information can help in improving the accuracy of sentiment classification and that there is a consistent need to comprehend and apply the multi-faceted concept of context in social data for real-time intelligence.

Table of contents

LIST OF ABBREVIATIONS

ANN	Artificial neural network
API	Application programming interface
Bi-LSTM	Bidirectional long short term memory
CA-LSTM	Context attention based long short-term memory
CBSA	Context-based sentiment analysis
CNN	Convolutional neural network
COAE	Context sensitive microblog sentiment classification dataset
CMU-MOSI	Carnegie Mellon University multimodal opinion-level sentiment intensity dataset
CRF	Conditional random forest
CV	Computer vision
CsHGCN	Context-specific heterogeneous graph convolutional network
DT	Decision tree
ET	Extra tree
GB	Gradient boosting
GIF	Graphics interchange format
GloVe	Global vectors
IMDb	Internet movie database
KNN	K- nearest neighbour
LDA	Latent dirichlet allocation
LR	Logistic regression
LSTM	Long short term memory
ML	Machine learning
MLP	Multi layer perceptron
MNB	Multinomial Naïve Bayes
NB	Naïve Bayes
NLP	Natural language processing
OCR	Optical character recognition
RBF	Radial basis function

R-CNN	Regions with convolutional neural networks
RF	Random forest
RNN	Recurrent neural network
RQ	Research question
SA	Sentiment analysis
Sarc-M	Sarcastic meme predictor
SLR	Systematic literature review
S-PLSA	Probabilistic latent semantic analysis
SVM	Support vector machine
H-SWN	Hindi-SentiWordNet
TF-IDF	Term frequency- inverse document frequency
WOM	Word-of-mouth

LIST OF FIGURE(S)

LIST OF TABLES(S)

CHAPTER 1

Introduction and outline

This chapter briefly introduces the research that was undertaken. Section 1.1 discusses the basic concepts and issues of sentiment analysis. The motivation to use 'context' in sentiment analysis is expounded & the scope of its usage is scrutinized for an improved performance of sentiment analysis. Section 1.2 discusses objective of research undertaken along with the statement of research question. Section 1.3 provides a brief description of the proposed models and their significance. Further, section 1.4 comprises of organization of thesis with summary of the chapter in section 1.5.

1.1 Introduction

Social media mining has proven valuable in numerous research areas as a pragmatic tool for public opinion extraction and analysis. Sentiment analysis (SA) addresses the dynamics of complex socio-affective applications that permeate intelligence and decision making in the sentient and solution-savvy social web. It encompasses investigation of both opinion and emotion within the content. Having started with simple polarity detection, contemporary SA has advanced to a more nuanced analysis of context, affect and emotion sensing. But detecting fine-grained sentiment in natural language is tricky even for humans, making its automated detection more complicated. Moreover, online opinions can be put forth in the form of text reviews or ratings, for a product as a whole, or each of its individual aspects. Multiple and lengthy reviews, usage of casual dialect with micro-text (wordplay, neologism and slang)[1], use of figurative language (sarcasm, irony)[2], multilingual content (code-mixed and code-switched)[3,4] and opinion spamming add challenges to the task of extracting opinions. Recently memes, GIFs, typo-graphic (artistic way of text representation), info-graphic (text embedded along with an image) visual content and edited videos dominate the social feeds. Consequently, the intra-modal modelling and inter-modal interactions between the textual, visual and acoustic components add to the linguistic challenges. Therefore, conceptualization and development of multi-faceted SA models to adequately capture observed opinion-sensitive information are imperative.

Typically, the key challenges [5] within the domain of SA either pertain to computational linguistics or to computational social science as given in figure 1.1.

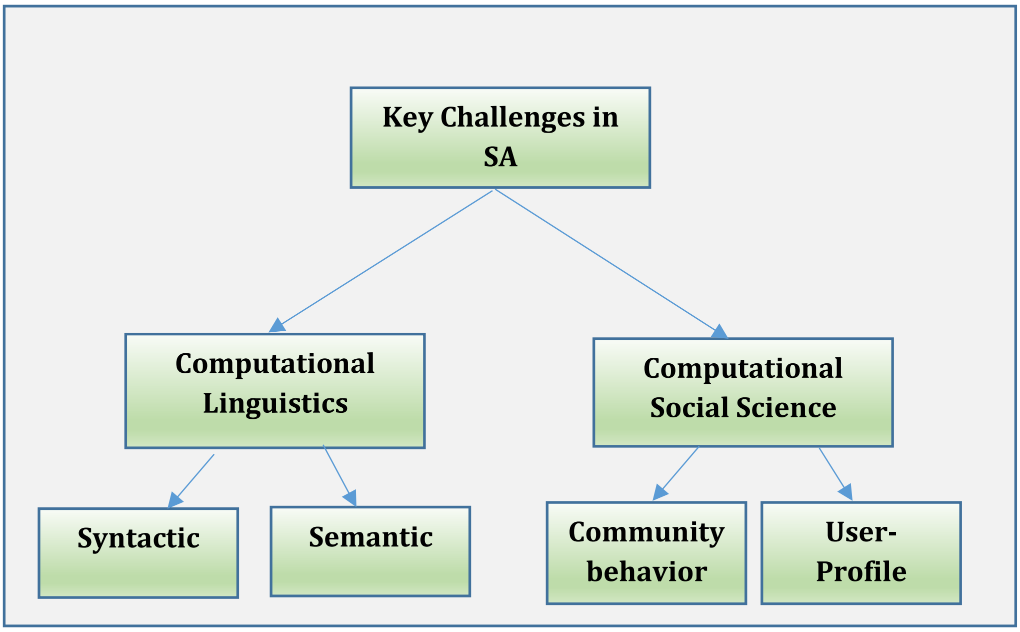


Fig.1.1. Key challenges in SA

1.1.1 Computational linguistics and natural language processing

Computational linguistics is essentially linguistics where computational tools and techniques are used to model human language processing. Natural language processing (NLP) [6] involves the usable knowledge of linguistics, artificial intelligence and computer science engineering in the tasks involving languages. SA is often described as a restricted NLP problem as it indispensably relies on the understanding of the human language and/or emotions expressed via textual or non-textual content. A language exists to transmit semantic information, and is composed of various key features, such as grammar and syntactics. These three general features often overlap extensively in language. Based on this, the primary linguistic issues within the domain of SA are further categorized into syntactic and semantic issues.

- ***Syntactic issues****:* The influence of social media sites and platforms on language is a true phenomenon. Though English is the 'lingua franca' (an intermediary language used by speakers of different language backgrounds) on social web, its ever-evolving nature leaves the language in a continuous state of alteration and regeneration. Slangs, colloquial words, non-standard abbreviations and mis-spelt words such as *'great'* as *'gr8'*, *'cool'* as *'cuuuul'* and *'ok*' as *'k'* introduce new vocabulary and at the same time alter the existing making it even more intricate to tap and analyse information using contemporary tools.

- ***Semantic issues**:* Essentially, "semantics" refers to understanding the meaning of the text. It is the study of the relation between linguistic expressions and their meanings. Social media texts are short and generally analyzed in an isolated manner and lack operational concern. A single word can have different polarity in different contexts. For instance, the word *'unpredictable'*, when used with movie plot, has a positive polarity but when used with a person's behaviour it's negative. Also, the use of figurative language (sarcasm, irony, humour, pun) which is a language that deviates from the conventional order and meaning in order to convey a complicated meaning further limits the capabilities of the conventional SA tools. Undeniably, the pragmatics that define the context or more precisely the way context can influence the understanding of linguistic utterances are missing.

  Further, baseline SA computes polarity strength but is incapable of capturing those opinions which can be shifted diagonally to the opposite direction given a particular scenario. Especially neutral polarity can shift to positive or negative. For example, the text *"What a surprise!", if* it is preceded by a text like *"The last ranking candidate admitted to the college topped the batch with distinction"* then the polarity shifts to positive. On the other hand, if preceded by a text like *"Topper of the class score zero in Mathematics",* the polarity shifts to negative. Accuracy in determining polarity must be improved. Apparently, there is a substantial shift of interest from mere estimation to reliable and more accurate polarities.

1.1.2 Computational social science

Computational social science involves studying the social and communication networks to analyze factors that influence human behavior and change social dynamics. They include concerns related to finding influential users, biased users, rumour- mongers, etc. People now are more interested in finding 'who' has written instead of only 'what' has been written. Polarity strength deviates from the assessed value in the presence of influence. For instance, for buying a car the customer will pay more attention to opinions posted by famous automobile experts rather than that of an ordinary person. Moreover, fine-grained SA further needs to understand and analyze aspects of social behaviors, interactions, and affective states based on observable actions. Affective computing, behavioural analysis, emotion recognition in plain posts and in conversations are various computational challenges that can improve the SA task within the socially connected ecosystem.

Understanding the pragmatics, emotion, cognition and behaviour are key to accurate SA. Ongoing research shows that some of the issues pertaining to natural language use can be resolved by adding extra information (i.e. context) to the process of SA. Semantics describes context-free meaning, coded into the content of the statement whereas the pragmatics describes nuanced meaning in a specific context. If the context supports this meaning, there is no conflict between the semantic meaning and the pragmatic meaning. If the context does not support the semantic meaning, then there would be conflict between the semantic and pragmatic meaning. Characteristically, context is defined as the parts of a written or spoken statement that precede or follow a specific word or passage, usually influencing its meaning or effect. For example, the text *“Unbelievable news!”*, if it is preceded by a text like *“An injured soldier scaled the Mount Everest in record time”* then the polarity shifts to positive. On the other hand, if preceded by a text like *“Two thousand people killed in an earthquake”,* the polarity shifts to negative. Hence, it is essential to capture the operational concern, that is the pragmatic meaning defined by *‘context’* for improving the SA task.

1.2 Research objectives

Automated text analytics has many promising practical applications and improvement in the same facilitates good decision making. The objective of this thesis has been to find ways to capture the contextual information about an entity of interest on social media and to use it for SA. It specifically aims at developing a CBSA model with improved accuracy of prediction and application of the same to deal with specialized sentiments, multimodality and multilinguality of data.

Statement of research question (RQ):

"How can the contextual information about an entity of interest on social media be captured and used for sentiment analysis?"

In order to improve accuracy of sentiment classification on social web, this unifying RQ was partitioned into the following three sub-questions. Each of these is addressed by this research:

- What kind of contextual information can be extracted from social media data?
- How can context be modelled, to accomplish the task of sentiment classification on social media?
- Which are the application areas where the SA of social data can be improved upon by use of context?

In accordance with the sub-questions, three research objectives were set as follows:

Research objective I –To comprehend the multi-faceted concept of context in social data.
Research objective II –To propose a novel model for CBSA.
Research objective III - Application of context for implicit SA of social data.

This research primarily aims to find out the types of contextual information which can be extracted from social media and can be applied to improve results of SA. In this direction, a multi-faceted conceptual framework for context has been built. It defines types of context that can be used in SA. The concept of context was then applied for building a model for contextual SA. In addition to this work, the contextual framework work dealing with single modality (textual data) has been extended to deal with multiple modality data. This conceptualization of 'context' was further applied for detection of specialized sentiment like sarcasm with improved accuracy. Sarcasm detection was carried out for both multiple modality data and multilingual data. This work presented a learning model for real-time sarcasm detection in Hinglish code-switch dataset. In this reference, table 1.1 represents the mapping between research objectives, RQs and research publication fulfilling the requirement of corresponding aim and query.

Table 1.1. Mapping of research objectives, questions and publications

ROs	RQs	Publication(s)
RO I	RQ1	Kumar A., Garg G. **(2019)**. *"Systematic Literature Review on Context-Based Sentiment Analysis in Social Multimedia."* Multimedia Tools and Applications. https://doi.org/10.1007/s11042-019-7346-5. Springer**. [SCIE- Impact Factor: 2.101]**
		Kumar, A., & Garg, G. **(2020)**. "*The Multi-faceted Concept of Context in Sentiment Analysis".* In Cognitive Informatics and Soft Computing (pp. 413-421). doi: 10.1007/978-981-15-1451-7_44. Springer, Singapore. **[Presented in April' 2019]**
RO II	RQ2	Kumar A., Garg G, (**2019**). *"Sentiment Analysis of Multimodal Twitter Data."* Multimedia Tools and Applications. 78, 24103–24119 (2019), https://doi.org/10.1007/s11042-019-7390-1, ISSN: 1380-7501. Springer. **[SCIE- Impact Factor: 2.101]**
		Kumar, A., Garg G. "*Fine Grain Sentiment Grading Of User-Generated Big Data Using Contextual Cues*" (**2020**). World Review of Entrepreneurship, Management and Sustainable Development. Inderscience. **[SCOPUS, In Press]**
RO III	RQ3	Kumar, A., Garg G. (**2019**). Sarc-M: Sarcasm Detection in Typo-graphic Memes (March 14, 2019). International Conference on Advances in Engineering Science Management & Technology (ICAESMT) – 2019. http://dx.doi.org/10.2139/ssrn.3384025. Elsevier [**Presented in March' 2019]**.
		Kumar A., Garg G, (**2019**). "Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets", Journal of Ambient Intelligence and Humanized Computing,1-16, https://doi.org/10.1007/s12652-019-01419-7, ISSN: 1868-5137. Springer **[SCIE- Impact Factor: 1.910]**

		Kumar A., Garg G. (**2020**) Sarcasm Detection Using Feature-Variant Learning Models. In Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering, vol 605. https://doi.org/10.1007/978-3-030-30577-2_61. Springer, Cham. **[Presented in July' 2019].**
		Jain DK, Kumar, A., Garg, G. (**2020**). "Sarcasm Detection in Mash-up Language using Soft-Attention based Bi-directional LSTM and feature-rich CNN", Applied Soft Computing, https://doi.org/10.1016/j.asoc.2020.106198, Elsevier ,ISSN: 1568-4946. **[SCIE-Impact Factor: 4.837].**

1.3 Proposed models for context based sentiment analysis

An attempt has been made to globally define and classify various type of context for usage in the future research from the contextual information that has been used till date in various aspects. A hybrid model incorporating contextual information for carrying SA based on machine learning (ML) and lexicon based techniques has been proposed. Also, this contextual framework work dealing with single modality (textual data) has been extended to deal with multiple modality data.

The concept of 'context in use' was explored further by using content-based local and global context to predict sarcasm in user-generated social textual data. Sarcasm is an implicit sentiment which is very difficult to capture in the absence of sufficient background knowledge.

An empirical analysis of number of computing techniques on different datasets like SemEval 2015 benchmark Twitter dataset, Reddit and random tweets was carried out for sarcasm detection on social media. Standard efficacy measures i.e. Precision, Recall & Accuracy have been used to analyze the performance of classifiers. Results validates that the proposed methodologies help in improving detection of specialized sentiment i,e, sarcasm. This was a preliminary work to understand the what, how and why of using context in sarcasm detection.

The contextual aspects were next experimented for multilingual sarcasm detection. This research proposes sarcasm detection using deep learning in code-switch tweets, specifically the mash-up of English with Indian native language, Hindi. More recently, as memes and GIFs dominate the social feeds; typo-graphic visual content has become a considerable element of social media. To deal with this type of multimodal data, the research presented in this work also attempts to detect sarcasm in typo-graphic memes. A model, Sarc-M (sarcastic meme predictor) to analyze sarcasm from visual language of Instagram memes has been proposed.

1.4 Organization of thesis

This section presents the organization of thesis. The thesis has been divided into eight chapters. Chapters' summary is as follows:

Chapter 1: This chapter briefly introduces the research that was undertaken. It discusses the basic concepts and issues of SA. The motivation to use 'context' in SA is expounded & the scope of its usage is scrutinized for an improved performance of SA followed by a brief discussion of the research problem undertaken. The unified RQ has been divided into sub questions leading towards certain research objectives. A mapping of RQs and objectives with the respective research publications is also included in the chapter. A short description notifying the significance of proposed framework has been illustrated in this chapter. Further, this chapter comprises of organization of thesis with summary of the chapter at the end.

Chapter 2: This chapter comprises of a state-of-art literature survey required to study and explore and analyze the existing work on the CBSA and to report gaps and future directions in the said research area. A brief description of the key terminologies such as SA, Contextual information, Subjectivity etc. has been done. The important practices followed for conducting such a review include surveys, narrative reviews, systematic literature review (SLR) and meta-analysis. The SLR is chosen for review in this research. Format of SLR given by Ketchenham and Charters was adopted for conducting the review process in this research. The review process was divided into six stages viz. formulation of RQs, search strategy, study selection, quality assessment, data extraction and data synthesis. Thereafter, the identified research gaps have been listed followed by the chapter summary.

Chapter 3: This chapter consists of the preliminary work done to understand the what, how and why of context using the context in SA. A multi-faceted concept of context has been described in detail. Chapter summary winds up the chapter.

Chapter 4: This chapter describe a model incorporating contextual information for carrying SA. An extension of the contextual framework work dealing with single modality to deal with multiple modality data has also been discussed in this chapter. Chapter summary winds up the chapter.

Chapter 5: This chapter describe the application of context in implicit SA. It presents how the 'context in use' can be applied to predict sarcasm in user-generated social data. This chapter primarily covers the description, methodology and findings of research objective 3. An empirical analysis of number of computing techniques on different datasets like SemEval 2015 benchmark Twitter dataset, Reddit and random tweets is presented. Standard efficacy measures i.e. Precision, Recall & Accuracy have been used to analyze the performance of classifiers. Validation of results showing improvement on using the proposed methodologies has been presented. Finally, the key points have been put into nutshell in chapter summary.

Chapter 6: A model for sarcasm detection using deep learning in code-switch tweets, specifically the mash-up of English with Indian native language, Hindi is discussed in detail. Detailed comparative analysis of results is also presented. Chapter summary winds up the chapter.

Chapter 7: This chapter presents the work done on sarcasm detection for multimodal data. A model to detect sarcasm in typo-graphic memes from the visual language of Instagram is discussed in detail in this chapter. Chapter summary sum up the key points of the chapter.

Chapter 8: This chapter recaps the research summary in conjunction with the limitations of the study. A thorough discussion of future scope and open areas of the research have been presented. And finally, conclusion concludes/wind up the thesis.

References line-up the details of citation sources used in the thesis.

1.5 Chapter summary

This chapter has put forward the groundwork for this thesis. It show the ropes of research problem, research objectives and the proposed solution framework. The need and motivation of the research area has been explained along with the organization of thesis.

CHAPTER 2

Systematic literature review

This chapter presents a systematic literature review on context-based sentiment analysis. Section 2.1 presents the introduction to SA process. Section 2.2 presents the stepwise review process followed for conducting systematic review. Section 2.3 presents overview of each selected study after quality assessment step in year wise reverse chronological order. Section 2.4 presents the analysis of the shortlisted studies to find solutions of the RQs. Section 2.5 lists the key observations and research gaps found on the basis of literature review conducted. Section 2.6 presents the chapter summary.

2.1 Introduction

With the advent of Web 2.0, usage of social media to gather opinions took a steep rise. The users across social media contributed a lot in generating huge volume of voluntarily disclosed information [7, 8]. The opinion seekers find this available information quite convenient and attractive. Such information is being used for their specific purpose be it business, social, educational or entertainment. However, the main difficulty lies in the analysis of such information for relating the same to the intended objective of the user. The key to meet such requirement lies in SA. SA is a process of systematic computational analysis of opinions, sentiments and emotions expressed in the text [9]. Such analysis is designed to detect the subjective information contained in the text and to discern the conviction of the author on an issue or the full text.

The term subjectivity is a keyword in the whole process of SA. It is described as the linguistic expression of somebody‘s opinions, sentiments, emotions, evaluations, beliefs and speculations [10]. The words opinion, sentiment, view and belief are used interchangeably though there are subtle differences among them [11,12]. The content under review can be original post of an author or reaction by others on such text. Such content may relate to a product, an issue, a person or an organization. Review sites, blogs, forums, and social network provides a good medium for analysis of sentiments.

SA involves analyzing a given piece of content by first pre-processing it to remove stop words, symbol etc. and then checking it for subjectivity. After getting opinionated content, polarity is determined either on lexicon based method or on ML based methods [13]. SA typically classifies the content into positive, negative and neutral. The polarity thus obtained has numerous applications for practical usage [12, 14]. As an example positive product reviews encourages the prospective buyers to go in for such a product, whereas the negative polarity will guide the producer to review future of product for corrective actions.

Although the SA had gained considerable popularity as a phenomenon for determining level of acceptability of a product, scheme, person or proposal but it showed limited performance when used for tasks where data was sparse, ironical, sarcastic, noisy, multilingual, ambiguous or consists of non-standard vocabulary [12, 14]. As a step forward to resolve these issues, some additional information was required. It was observed through various pertinent literature that context provides such additional information. Traditional SA methods were more or less domain independent. SA makes use of knowledge. Knowledge is domain dependent and validity of facts change along with context switches [15]. Hence, aforesaid problems of SA can be resolved by applying appropriate context to SA. Context-based Sentiment Analysis (CBSA) is process of applying context to traditional SA aiming to improvise accuracy of results. Different approaches have been used for CBSA ranging from using contextualized lexicon to hybridization of lexicons and ML techniques in respect to context. Different type of context that is linguistic, user profile, social, spatial, temporal etc. can be applied in singleton as well as in combination to achieve better results.

To explore the usage of context in SA so that the existing trends and scope for further research could be identified, a thorough review was required. The practical usage of SA began around the year 2001. Mike Chang et al. in 2001[16] used SA for extracting market sentiment from stock message boards. Turney [17] and Lee [18] in 2002 used unsupervised classification and ML techniques respectively for SA. Though the popular social media LinkedIn 2002, Myspace 2003, and Facebook 2004 became means of influencing mass opinion but the real spurt came with the advent of Twitter in 2006[19]. The important practices followed for conducting reviews include surveys, narrative reviews, SLR and meta-analysis. We chose the SLR for review in our work. An SLR seeks to systematically search for, appraise and synthesize research evidence often adhering to the guidelines on the conduct of review [20].

2.2 Review process

In order to comprehend a state-of-art within the area of CBSA, a SLR on the basis of format by Ketchenham and Charters [20] was conducted. The review process was divided into six stages namely formulation of RQs, search strategy, study selection, quality assessment, data extraction and data synthesis. The flowchart of the review process is given in figure 2.1.

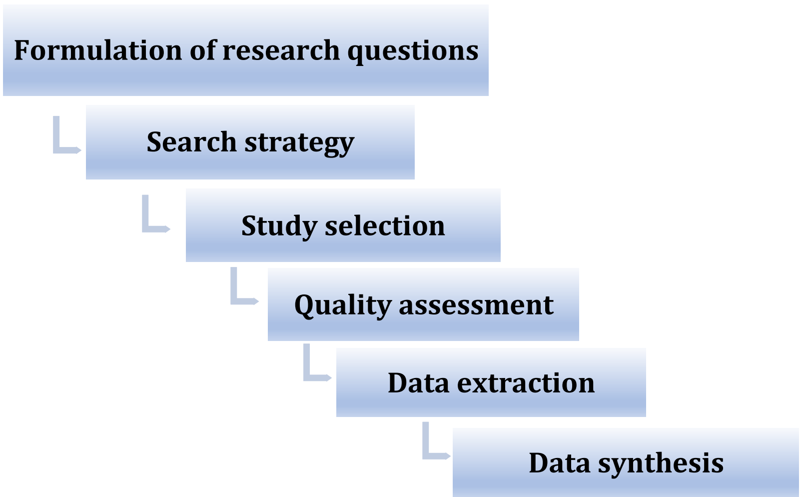


Fig.2.1. Stages of review process

2.2.1 Research question formulation

In order to collect and collate adequate research evidences from the available studies, in the domain of CBSA, the following RQ were framed:

RQ1: What is the need of using '*context*' in SA?

RQ2: What are the types of '*context'* which have been used in SA?

RQ3: Which are the techniques used for CBSA?

RQ4: Which are the social media in which CBSA have been applied?

2.2.2 Search strategy

After framing the RQs, a well thought list of key word and phrases was drawn up. These keywords assisted in locating all the relevant studies which are available in identified digital portals. The following terms were extracted from RQs: context, SA, opinion mining and social media. These search terms defined the initial search strings. The search based on the synonyms of the keywords was also carried out. For example, twitter, cue, hint etc. were also used in search. Further, OR/AND Boolean expression and wild card based search was used for expanding or narrowing the sweep of the search, as required.

Digital libraries/e-portals/online bibliographical databases are globally accessible, appropriate and extensive place to search relevant studies. These sources provide the details of the articles published in various Journals and Conferences. In this SLR six digital portals namely IEEE Explore, Elsevier, ACM, Taylor and Francis, Wiley digital online and Springer were used for searching of appropriate studies. The search indices were limited to the meta-data i.e. Title, Abstract, and Keywords excepting Springer where only the title search was performed. Cross citation were also tapped as additional studies to avoid missing of any important study. The search was restricted to a time span of little over a decade commencing 2006 and concluding in December-2018.

2.2.3 Study selection

In order to conduct a focused study, the selection of papers was restricted by applying the following 'inclusion- exclusion' criteria:

- *Inclusion criteria:*
  - ❖ Primary studies on CBSA
  - ❖ Studies proposing different definitions and types of the term '*context*' used in SA
  - ❖ Surveys & secondary studies on CBSA
  - ❖ Comparative studies on CBSA:
    - ➢ Comparing techniques used in CBSA
    - ➢ Comparing CBSA with non-CBSA
  - ❖ Benchmark evaluation studies with challenges in the form on shared tasks (SemEval)
  - ❖ Studies published only in journals and cross-referenced conferences.
- *Exclusion Criteria:*
  - ❖ Studies belonging to non-CBSA.
  - ❖ Studies on CBSA in other than English language data sources.

2.2.4 Quality assessment

Quality assessment was used to establish a 'quality threshold' below which the studies were excluded. It was used to distinguish the studies in terms of overall contribution and to gain an understanding of the relative strengths and weaknesses of the body of evidence [21]. In this SLR, a quality assessment strategy was adopted depending upon the number of weighted parameters as follows:

- Novelty: What was the level of newness in the proposed idea or it was mere enhancement in any existing version?
- Content & Analysis: Was the content technically sound and well supported by evidence and theories, comparative merits over state of art methods provided or not?
- Results: Was the result well presented, used benchmark data set or not, clear outcome obtained or not?

Each study was given a score out of 10. The score was split as per the following values: 2 for novelty, 4 for content & analysis and rest 4 for results in which 2 were for data set and 2 for evaluation criteria. Table 2.1 depicts the quality assessment of studies

Table 2.1. Quality assessment

Quality level	Number of studies	Percentage
Outstanding (score > 9)	9	10.11 %
Excellent (8 < score <=9)	11	12.35%
Good (6.5< score <=8)	8	8.98 %
Average (5 <= score <=6.5)	9	10.11%
Poor (score <5)	52	58.42%

Application of quality assessment criterion has led to reduction in number of papers finally included in the review. Studies rated as poor or below average were not included in the final selection in the review.

2.2.5 Data extraction

The aim of this step was to extract information from the selected studies and find out which RQ was answered by which study. A mapping exercise was undertaken for the purpose. Details about author, year, publisher, technique, dataset, context, description and result of proposed method were extracted from each of the selected study.

2.2.6 Data synthesis

Data Synthesis involved the collation, combination and summary of the findings of individual studies included in the systematic review [21]. The information obtained from the data extraction stage was combined, summarized and interpreted using different visualization methods like tables, graph and charts.

In brief, the research process in this SLR started by applying search terms on six e-databases identifying 573 papers. After removing redundant studies out of 573, we were left with 279 studies. On applying inclusion and exclusion criteria only 89 studies were found suitable for further analysis. When put to quality assessment criterion, only 30 studies were found suitable. Another set of 12 studies were derived from reference using cross referencing. These additional studies were also put to quality assessment criterion, to get the papers relevant to current SLR. This led to selection of 7 papers. For cross referenced papers exclusion criteria of being only from journal was dropped to cover the subject matter of this review in sufficient breadth and depth. So, total of 37(30 +7) papers were identified to analyze and study further. Lastly in the data extraction and synthesis step, information was retrieved from 37 studies and presented in tabular format, graphs and charts. Final output of review was summarized as to answer all RQ addressed, finding research gaps and future directions. Following figure 2.2 shows the review process adopted in this SLR:

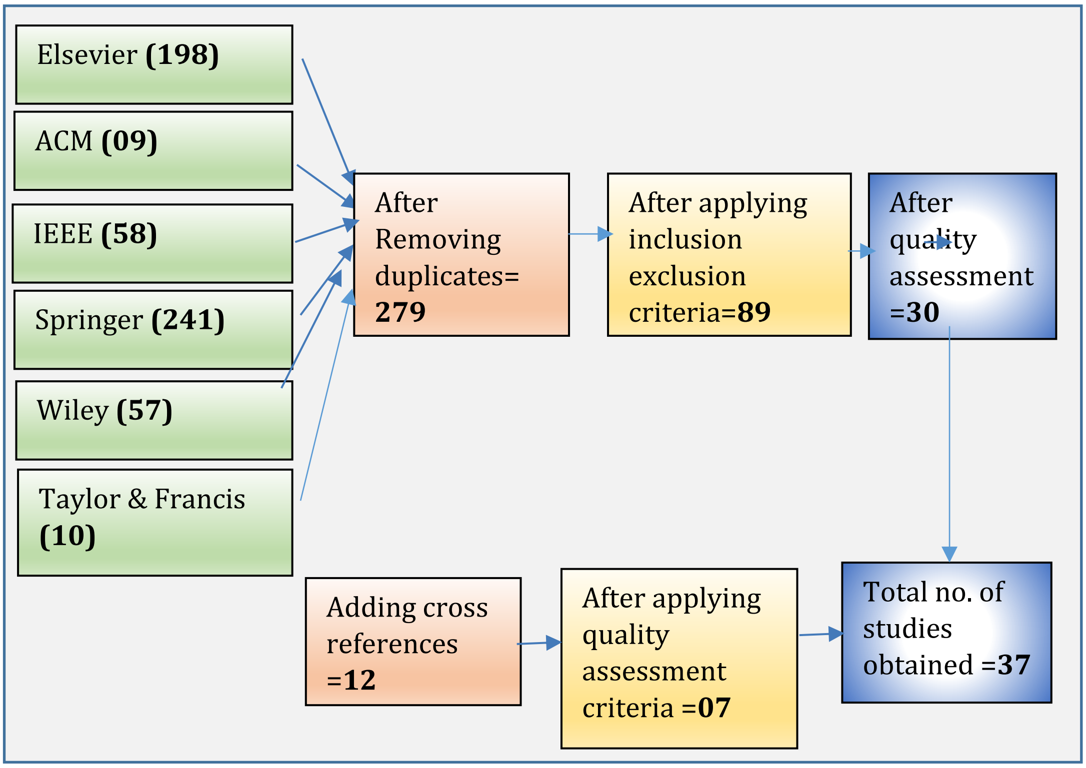


Fig.2.2. Detailed review process

2.3 Literature survey

A brief overview of each selected study after quality assessment step is presented in this section in year wise reverse chronological order. In each year there has been substantial work done, details of which are summarized in the table 2.2.

In 2018, Han, H et al. [22] proposed a novel model based on neural network to capture context information for the sentiment classification task. The authors used local semantic information of a word context to improve sentence representation. They also provided comparative analysis of results for three different datasets viz. IMDb, Yelp 2014, Yelp 2013. Majumder et al. [23] proposed a novel feature fusion strategy for multimodal SA. The authors also used context modelling and showed an improvement in classification results by margin of 1-2% for all the modalities combinations. Sheik et al. [24] worked on extracting most influential sentiment for topics and used Twitter rank algorithm for finding influential users for a topic. They also used SentiCircle for capturing entity level sentiment and showed that their methodology gives more accurate results as compared to supervised or lexicon based method. In [25], authors modelled microblog conversation as sequences on the basis of time. They developed a context attention based Long Short

Term Memory(LSTM) network and incorporated attention mechanism into it to get sentiment weights.

In 2017, Olga Vechtomova [26] proposed a novel method named PolaritySim for disambiguating context dependent polarity of words. Readily available reviews with numerical ratings were demonstrated to be effective positive and negative reference corpora for determining word level contextual polarity. An information retrieval approach was used wherein positive and negative corpora were treated as "documents" and a specific instance of word in a sentence was treated as a "query". Similarity between query and documents were then calculated and the polarity of best matching document was assigned to query. The selected studies in 2017 include seven more studies published in the same year. Chihli Hung [27] proposed a novel SA approach which uses contextual lexicons and preference vector modelling to improve quality classification for word of mouth (WOM) or opinionated text. Three elements of WOM i.e. a rating score, a concept category and context were used for building contextual lexicons. Shuyuan Deng et al. [28], used dependency features for exploring contextual information. Dependencies were calculated from relations among adjacent as well as non-adjacent words. They incorporated this contextual information into supervised classification (Support Vector Machine (SVM), Naïve Bayes (NB)) technique and showed that resultant system preformed 1% better than baseline methods. Wantao & Tao Liu [29], proposed a new ontology based method which helped in finding opinion feature related to different contextual environment. Annotations were done using context – distance & co-occurrence number. Proteg 5.0 & OWL 2, were used for ontology building. Hassan Saif et al. [30], used lexicon based approach for sentiment classification. They proposed a generalized method for adapting lexicon to given context. In comparison to baseline techniques, an increase of 3.4% in accuracy and 2.4% in F1 score was achieved. YunxiaoZhou et al. [31], used a set of features viz. NLP, domain & word embedding for sentiment classification via supervised method. Specifically in domain, they used tweet metadata and user metadata to improve the result. Jimenez-Zafra et al. in [32] took user information from their timeline and used it to build a user model. This user model along with SVM was then used for classifying the tweets. E.Fersini et al. [33], developed a model which uses homophily and constructuralism for defining approval network. User's behavior on social network and user's network were used to make a graph. This graph represents user's relationship and textual content. The added information in the form of approval network improves the effectiveness of proposed model.

From amongst the studies published in the year 2016, seven studies were selected for inclusion in this SLR. Aminu et al. [34] built smart SA framework by incorporating contextual information into lexicon based Model. Context was captured from both local and global level. The dependence of contextual polarity on domain knowledge was explored. Detailed results of comparative analysis presented, showed the efficiency of proposed methodology. Hassan Saif [35] captured contextual semantics to refine pre assigned polarity & strength of sentiments. Three data sets STS-gold (Standford Twitter Sentiment Corpus), HCR (Health Care Reform) and OMD (Obama-McCain Debate) were used with lexicon Based approach. Improvement in accuracy and F-measure was achieved using proposed idea. Matthijs Meire et al. [36], demonstrated the added value of auxiliary data in SA. Pre-post and after-post information was obtained and applied to baseline SA methods in three different ways. Experimental setup showed improvement in predictions specifically for informal messages. Preslav Nakov et al. [37], discussed the concept of contextual polarity disambiguation by taking context around target term into consideration. Dependency feature, information from neighboring terms, topical information etc. were used to capture contextual polarity. Improvement in results were shown for twitter data. Will Frankenstein et al. [38], discussed the role of context in SA. Merits of application of context in changing previous ratings and evaluations were discussed in detail. FangzhaoWu [39] built a framework for structured microblog sentiment classification using social context. User, friend and topical contexts were applied individually as well as in combination. Proposed framework performed better than content based methods. Among the different ways of application of context, the performance of combined application of two or more contexts was found better than the application of singleton. Also, usage of 'friend' context improved the performance more than the usage of 'user' context. Rui Gaspar et al. [40], took event related context into consideration and discussed the qualitative SA. Event related context can act as sentiment modifier. Findings about how people react to a critical situation, can provide means to improve simple positive-negative classification.

Similarly, four studies of the year 2015 have been included in this SLR. Francesco et al. [41] in 2015, proposed a method to predict popularity score of images by using visual, context and user features. Tag type, tag domain, description were used as contextual features. Also, showed that usage of context improves the prediction power of underlying method. In [42], authors presented a model for real time content analysis. They focused on finding disruption and disorder level occurring during public events using SA. Evidence based functions and relationships between sentences were used as key ideas. Authors in [43], developed a method to improve accuracy in SA by taking 'popularity at a location' and 'gender information' into consideration while classifying sentiments. They used artificial intelligence tool for extracting gender information from twitter data. Mauro Dragoni et al. in [44], used fuzzy logic for modelling concept polarities based on particular domain. They concluded that worst performances were obtained on domains where only a small number of instances were available in the dataset. Diego Roforgiato et al. [45], proposed new version of "SENTILO" which is a sentic computing system for SA. Two types of scores were calculated, one for identified topic and the other for overall sentence. The information regarding opinion holder, topics, situations and events were extracted and used in the process of SA.

From those published in the year 2014, seven studies were included in this review. In [46], sentic computing framework was used in conjunction with dependency based rules. Dependency rules were used to find contextual role of concepts. Author showed that proposed method improved the accuracy of polarity detection. Yung-Ming Li et al.in [47], proposed a personalized social context endorsement mechanism in which increase in click through rate was achieved using social influence. Social context was applied to identify target receivers for an advertisement. Context embellishment was used by authors to improve acceptance level of an advertisement by the target user. They also tried to capture social emotional context through emoticons. Comparative analysis of proposed model performance against number of benchmark methods was also presented. Raymond et al. [48], extracted context sensitive sentiments from consumer and applied the same in semi supervised learning method. Proposed model computed the polarities from product ontologies (provides context) and context free lexicons. Bishan Yang et al. in [49], used local and global contextual information for sentiment classification. They took lexical knowledge as local context and discourse knowledge as global context. They showed that by adding contextual knowledge to conditional random forest model (CRF), better accuracy was achieved in comparison to existing sentence level classification methods. In [50], the author incorporated a social influence factor into SVM, NB, Max Entropy and Artificial Neural Network(ANN) for classification of tweets.

Two different datasets were tested for all of the four techniques. Re-tweet count was taken as the added influence factor and improvement in accuracy level were shown. In [51], author explained that research on SA over twitter, treats SA to be a function of incoming tweets. In their proposed method, they filtered the tweets according to the topic and tried to capture context by sequentially tagging the previous tweets. They also used user profile partially in their work. They concluded that there is future scope for deeper study on usage of user profiles specifically on reputation, authority and influence of user. Authors in [52], proposed a new method for SA based upon appraisal theory. Sentiments were identified with respect to target of a post which can be a person, an object or an event. Appraisal expressions giving information about attitude, engagement, graduation etc. were used for finding context.

In 2013, Fuji Ren et al. [53], devised a framework named 'ScTcMf' for predicting the implicit opinion towards a topic. The authors mathematically modeled the social context captured by opinion homophily and topical context. Social and topical context when added to traditional approach for SA showed good results measured by Recall andF1 score. For assignment of label to user-topic opinions senti-strength was used. The authors in another study [54] discussed the challenges faced at each step in the process of development of a corpora for SA specifically for irony detection. Authors stated that textual and dialogical context helps in the detection of irony. Albert Weichselbraum et al. [55], collected context terms for each ambiguous term for resolving ambiguity. They built a general contextual sentiment lexicon by combining contextualized lexicons of multiple corpora .Overall improvement was shown when contextualized lexicons were used in SA. Yang Liu et al. [56] devised an adaptive SA method. Authors captured the sentiment changes from latest available reviews and revised the sentiment polarity based upon it. The out of date data was discarded on the basis of temporal changes in the data. When tested on movie review data, authors showed that results were more accurate as compared to those obtained from non-adaptive methodology.

In 2012, Fotis Aisopos et al. [57] devised a method by combining CBSA and content based SA. The authors used social graph connections for capturing mood expressed in each message. They also discussed the role of polarity ratio in context driven SA. A social context model by taking various parameters like followers, followees, topic URL etc. into considerations was also build.

In 2009, Lipika Dey et al. [58] extracted opinions from noisy data using domain knowledge. By using neighboring words as context, spelling corrections were done. On the basis of context, disambiguation of word as opinionated or modifier was done. Authors used customer feedback and consumer review data to test the proposed technique and achieved 89% accuracy.

The techniques, authors, data set used, details and scope of proposed method, type of context used in the studies referred to above are presented in appropriate form in table 2.2 below.

Table 2.2. Year-wise summary of selected studies

S. No.	Author, Publisher, Year	Techniques	Dataset	Context type	Details & scope
1	Han et al. [22] Springer, 2018	Convolutional neural network (CNN), LSTM	IMDb, Yelp 2014, Yelp 2013	Local (correlation between words and sentences)	• A novel model based on neural network to capture context information for sentiment classification • The local semantic information of a word context was used to improve sentence representation
2	Majumder et al. [23], Elsevier, 2018	Recurrent Neural Network (RNN)	CMU-MOSI	Multiple utterances of video clips	• A novel feature fusion strategy was proposed for multimodal SA • Context modelling was used
3.	Sheik et al. [24], Springer, 2018	Latent Dirichlet allocation (LDA), SentiCircle	Twitter	Topical, social (number of followers, followee)	• Worked on extracting most influential sentiment for topics • Twitter rank algorithm was used to find influential users for a topic.

					• Entity level sentiment were captured using senticircle
4	Feng et al [25] Springer,2018	LSTM	*Context* sensitive microblog sentiment classification dataset (COAE-2015)	Textual (order of words in tweet, order of tweets in sequence)	• Modelled microblog conversation as sequences on the basis of time. • Developed context attention based LSTM network. • Incorporated attention mechanism into LSTM to get sentiment weights.
5	Vechtomova [26] Elsevier, 2017	Information retrieval model	Amazon corpus, SemEval ABSA 2016 test dataset, restaurant review data	Local(lexico-syntactic)	• A novel method named PolaritySim was proposed. • Task of resolving polarity is taken up as an information retrieval problem. • Lexico-syntactic features were generated from a dependency parse graph of sentences. • Comparison with SVM and MNB (Multinomial Naïve Bayes) classifiers were presented
6	Hung [27] Elsevier, 2017	Lexicon based approach, preference vector	IMDb, hotel.com	Domain	• Contextualized lexicon were used to capture relation between tokens & their

		modelling, SVM (linear & RBF), decision tree (DT), NB			associated concepts and categories. • Contextual lexicons were combined with preference vector modelling to improve word of mouth quality classification. • For building WOM trained classifiers SVM linear kernel, SVM with RBF kernel, J48 DT and NB were used
7	Deng et al. [28] ACM, 2017	Supervised Classification (SVM, NB)	Stockwits messages	Co-occurrence of words (Adjacent, Non-Adjacent words)	• Taken up the issue of polysemy of words • Dependency features were taken as sentiment indicators • Dependencies explored contextual information by using relations among adjacent as well as non-adjacent words
8	Tao et al. [29] Taylor & Francis, 2017	Ontologies (Proteg 5.0, Owl 2)	Domains of Weather, Book and Shopping	Domain	• An ontology "DEMLonto" based on six basic emotions to help users to share existing information was proposed "DEMLonto" helps in finding the opinion features related to different contextual environments

					• Stated that vocabulary is the basis of SA • Entities are annotated at DOM level as derived from context-distance and co-occurrence number
9	Saif et al. [30] IOS Press, 2017	Lexicon Based Approach	Twitter data	Co-occurrence of words	• A general method for adaptation of usual lexicons to given contexts • Drawn comparison of proposed contextualized lexicon adaptation method and baseline techniques.
10	Zhou et al. [31] ACL, 2017	Supervised ML	Twitter data	NLP features, domain, user profile	• Uses number of NLP, domain & word-embedding features for classification • Tweet metadata and user metadata were also included in domain features.
11	Zafra et al. [32] ACL, 2017	SVM	Twitter data	User profile	• User information was used in SA • User model was obtained from his timeline • Based upon the accuracy level, one among the user model or the general SVM

					model is applied on test set
12	Fersini et al. [33] Springer, 2017	Semi supervised & unsupervised classification	Twitter data	Social (User behaviour on social network, user social network)	• A new method of SA based on usage of Approval network was proposed • Approval network was used for modelling homophily & constructuralism • HDAG was used to model user relationship & textual contents • Comparison with NB, SVM, CRF, Bayesian, Max Entropy technique was done.
13	Muhammad et al. [34] Elsevier, 2016	Lexicon Based Sentiment Classification	Twitter, Digg, MySpace	Local Global	• A new framework "Smart SA" was proposed • Worked on the idea that term polarity depends on the domain in which it appears (Domain context) • Captured Context in two ways: ➢Local context – neighboring terms ➢Global context – Domain specific

					• Hybrid lexicon was developed by adding local and global context to general lexicons
14	Saif et al. [35] Elsevier, 2016	Lexicon Based approach	Twitter data	Co-occurrence of words	• New Technique Senticircle was proposed • Captured semantics of words from co-occurrence patterns of words to update pre-assigned strength and polarity of sentiments • Performed analysis on both entity and tweet level • Used three different lexicons and data sets • Constructed STS-Gold data set
15	Meire et al. [36] Elsevier, 2016	Random Forest (RF), SVM	Facebook	User Profile, Textual	• By incorporating extra information (before post and after post information) in baseline SA, a new model was developed • Make use of 1) Leading information: User Profile, Previous Posts 2) Lagging information: Future Likes, comments • Three variants of model were

					developed: 1) Model using past information. 2) Model using present and past information. 3) Model using present, past and future information • Explore the relationship between main predictors with sentiment of posts
16	Nakov et al. [37] Springer, 2016	SVM, Lexicon based approach	Twitter data, NUS SMS corpus, Sentence from Live journal	Co-occurrence of words (Bigram near target terms and target phrases, Parse features)	• Concept of contextual polarity disambiguation was discussed • Number of techniques for contextual polarity disambiguation were applied: 1) Extraction of unigram and bigram within 4 words on either side of target term. 2) Use of dependency parse features. 3) Features from neighboring target phrases • Classification of message polarity on the basis of topic was done
17	Frankenstein et al. [38]	Dictionary based method	Twitter	Textual (In relation with	• Discus role of context in SA

	Springer, 2016			original post)	• Comparison between contextual and non-contextual approaches for same dataset • Implication of applying context for existing tools was discussed.
18	Fangzhoao et al. [39] Elsevier, 2016	Logistic Regression (LR)	Twitter data	Social (user & friend)	• Structured micro blogs sentiment classification (SMSC) framework was proposed • Social Context was added (as graph structure) to base line SA methods to improve accuracy • Types of social context used: topic, user and friend • Addition of individual context type as well as combination of context was tested • Comparison with SVM, NB, LR, LPROP and SANT was done.
19	Gaspar et al. [40] Elsevier, 2016	Human Based Qualitative Analysis	Twitter	Event Related	• Studies importance of qualitative SA • Added a qualitative analysis layer to computer-based SA

					• Event related context-based analysis was discussed
20	Gelli et al. [41] ACM, 2015	CNN	VSO Flickr dataset	Domain, Textual, Topical	• Proposed a method to predict a popularity score of images by using visual sentiment features, 3 new context features and user features (author data) • Relates the extraction of entities from image context to the nature of text (tags, textual descriptions) • Tag type, tag domain, description are taken as contextual features
21	Jurek et al. [42] Springer, 2015	Lexicon Based Approach	Stanford test Twitter corpus	Event related, Local (Negations, intensifications)	• A method for performing real time content analysis was proposed • SA was done for finding level of disruption and disorder during public events • Evidence based combination function and normalization was used to calculate the strength of sentiment

					• Combination function models the relationship between sentences by taking into account the number of non-neutral words and the value of sentiment they have
22	Hridoy et al.[43] Springer, 2015	Sentiword Net, NamSor (AI tool)	Twitter	Location of tweet, User profile (gender of user)	• Uses popularity at a location & gender information from data for improving accuracy in SA • For gender identification NamSor (AI tool) was used
23	Dragoni et al. [44] Springer, 2015	Fuzzy logic	Blitzer (Amazon data)	Domain	• Fuzzy logic was used for modelling concept polarities based on particular domain • Knowledge graph was made by using WordNet and Sentic net
24	Diego et al. [45] Springer, 2015	Sentic computing	Hotel review data from tripadvisor.com	Topical, event related	• A newer version of Sentilo (Sentic computing system for SA) was proposed • For an opinion sentence, sentiment score for each identified topic and

					overall sentence was calculated • Sentilo tries to find opinion holder, topics, situations, events and uses this information in SA
25	Poria et al. [46] Elsevier, 2014	Sentic Computing	Movie Reviews, Blitzer (Amazon data)	Discours-e structure informat-ion	• Proposed technique incorporates dependency-based rules into Sentic computing framework • Dependency relations used in sentences helps in capturing contextual role of concepts • Knowledge context of concepts leads to improvement in polarity
26	Yung et al. [47] Taylor & Francis, 2014	Dictionary based approach	Facebook	Social	• A personalized social context endorsement method was proposed • Social influence was used to find target recipients • SA was used to find the positive comments about products • To classify polarity distant weighted count was used

					• Emoticons were used to capture socio emotional context
27	Lau et al. [48] Elsevier, 2014	Ontologies, LDA based topic modelling, SVM	Comments from Epinions.com, IMDb.com, tripadvisor.com	Topical	• Proposed a model for automatically extracting fuzzy product ontologies (which gives contextual information) • Captured social intelligence from consumer comments • Developed an ontology based product review miner for evaluation of proposed technique • The context sensitive polarity was determined by using product ontologies and context free lexicons
28	Yang et al. [49] ACL, 2014	CRF model	Customer review, multi-domain Amazon data	Local, Global	• Sentence level classification was done using local and global context • Local context is intra sentence (lexical) information and Global context is inter sentence (domain) information

					• Uses posterior regularization for SA
29	Anjaria et al. [50] Springer, 2014	SVM, NB, ANN, Maximum Entropy	Twitter (US Election and Karnataka State Election)	Social	• Used SVM, NB, ANN and Maximum Entropy for classification of Twitter data. In all these novel influence factor (re tweet count) was combined with SA to improve prediction accuracy
30	Andrea et al. [51] International conference on computational linguistics, 2014	SVM	SemEval 2013 data set (Twitter data)	Topical, User profile	• Topical context was captured by tagging sequence of previous tweets • User profile was used to capture static view of context • Topical &user profile context was used in SA
31	Korenek et al. [52] Springer, 2014	SVM	ST02 dataset (Twitter data)	Appraisal expression	• A new method for SA based on appraisal theory was proposed • Sentiments were identified with respect to target of microblog post • Method is independent from topics of microblog

32	Ren et al. [53] IEEE ,2013	Matrix Factorization	Twitter Data (Active user profile and their tweets)	Topical, Social	• Uses concept of homophily • Used user's opinion consistency on content related topics • Knowledge from is used in conjunction with social and topical context to predict unknown user topic opinion • New framework obtained was named as ScTcMf
33	Bosco et al [54] IEEE, 2013	Manual Annotation	Twitter	Textual and Dialogical	• Discuss challenges faced at each step in the process of development of corpora for SA • Focused on manual development of corpora for Irony detection (textual and dialogical context helps) • Tests conducted for hypothesis that "Ironic expressions are polarity reversers" • Usage of built corpora in emotion detection was shown

34	Weichselbraun et al. [55] IEEE, 2013	Lexicon based approach	Product review data (Amazon.com), Hotel Review data (Tripadvisor.com), IMDb	Concurrence of words (negation, intensifier)	• Contextual sentiment lexicon were made by collecting context terms for each ambiguous term and use them for solving ambiguity. • Contextualizes lexicon of multiple corpora were combined to form generic contextual sentiment lexicon • Tried to replace contextual sentiment lexicon with knowledge bases
35	Liu et al. [56] Springer, 2013	S-PLSA (quasi Bayesian model)	IMDb	Temporal	• Proposed an adaptive SA method • Uses temporal data • Discards he out-of-date data • Captured sentiment changes from newly available reviews
36	Aisopos et al. [57] ACM, 2012	NB, SVM, C4.5	Twitter	Social, temporal	• Comparative analysis of content and context SA • Suggested a new metrics “Polarity Ratio” and its usage in context driven SA • Build a “Social context model” by taking number of authors

					tweets, followees, reciprocal friends, number of topic mentioned and URLs in account
37	Dey et al. [58] Springer, 2009	Human assisted knowledge base approach	Customer feedback from www.indiacar.com, customer review data from www.cs.uic.edu	Domain, Local	• Opinion were extracted from noisy data using domain knowledge • Context-based spelling correction was used • Disambiguation of a word as opinionated or modifier was done on the basis of context

2.4 Analysis of selected studies

In this section, analysis of the selected studies obtained in section 2.3 along with solutions of the RQs defined in section 2.2 of SLR are presented. To analyze the growth of research in this field, we examined studies in year-wise manner.

Following figure 2.3 represents year wise distribution of studies selected in the SLR.

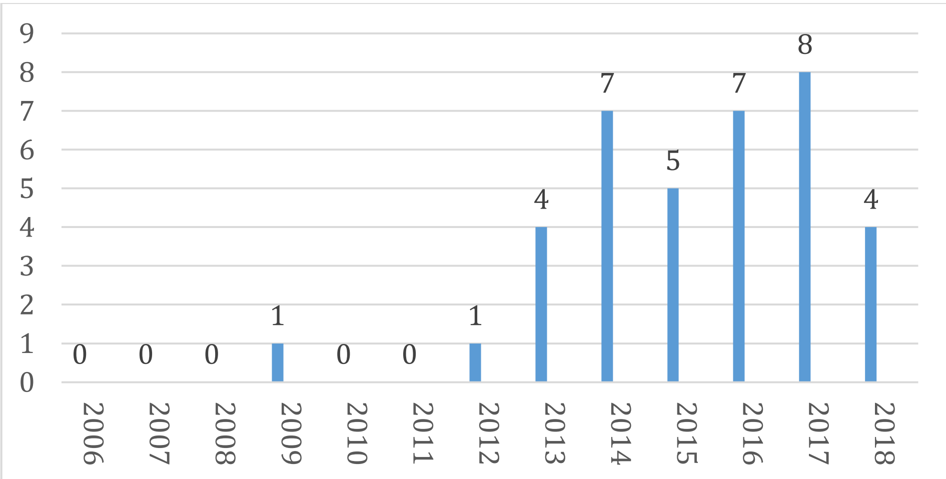


Fig.2.3. Year wise distribution of studies

After the scrutiny of journals in which the selected studies were published, the major work was found in publication by Springer (Table 2.3).

Table 2.3. Distribution of selected papers w.r.t. journals

S. No.	Publisher	No. Of Papers
1	Springer	14
2	ACM	03
3	Elsevier	10
4	IEEE	03
5	Taylor & Francis	02
6	Wiley	00
7	Others (special series /cross referenced)	05
Total no. of papers =		37

RQ1 intends to find out the need of using context in SA. Following observations were made based on this SLR:

- Polysemy of single words is difficult to handle. Meaning of words can vary greatly with respect to context in which they are used. For example, the word 'bank' has two meanings viz 'river bank' or 'financial institution' in the sentence "I am waiting for you near the bank".
- Microblog Messages are short & hence ambiguous. This makes their classification a difficult task when studied in isolation. Topical context or discourse information can resolve the ambiguity. For example, the text "Stop yourself" is ambiguous. When preceded by an advice, it indicates positive polarity and if by a criticism then the polarity becomes negative.
- Sometimes Text does not provide sufficient information about the hidden sentiment i.e. sentiments are implicit. Knowledge of discourse can resolve these issues. For example, in the text "I made sincere efforts. Now I accept the outcome without any grudge" has implicit sentiment.

- Sentiment polarities are context dependent, so context is important. Context change can vary the polarity. For example, the text "It will rain tomorrow". If this text is read in the context of crops, it is positive while if used in context of cricket match, it becomes negative.
- Neutral polarity shifted to either positive or negative in respect to knowledge of context which is otherwise difficult to classify in its absence. For example, "unpredictable" is a text with neutral polarity. If used in the context of a person's behavior, it carries negative polarity otherwise when related to a movie plot, it becomes positive.
- To detect and classify, irony and sarcasm some extra information is required as the sentiment is not explicit in two types of the language construct. For example, in the sentence "What a wonderful host!" sentiments are not explicit. The background information about the event hosted only can resolve the hidden sentiment.
- Ill-formed syntax &nonstandard vocabulary usage makes traditional methods show poor performance. Knowledge of context of content being reviewed can solve the issues. For example, the text "I will come ASAP" contains nonstandard usage of vocabulary. If preceded by a request for urgent arrival then the usage of ASAP is understood to be 'as soon as possible'.
- Sentiment change when subjected to different event & stressful situation. For example, financial aid to poor people is a positive gesture in normal times but negative if done around election time.
- Sentiment of a user towards a topic changes over a time so temporal context factor must be taken into consideration. For example, a person may be favorably disposed towards a practice and after some time he may be neutral or bitter critic of the same.

RQ 2 intends to find out the ways in which the 'context' has been used in SA. The literature survey presented in table 2.2 of section 2.3 summarize the different type of contexts used in studies included in this review.

Table 2.4. Mapping of type of context used & studies using them

S. No.	Context name	Definition referred	Study number
1.	Social	Social network information (linked users)	[53],[33]
		Connection between two messages. Two subtypes of social context: 1)User context : connection between messages brought by same author 2)Friend context: connection brought by relation between users	[39]
		1)Data about "likes" and friends of users 2)Social influence of users 3)Socio- emotional context using emoticons	[47], [24]
		Any indication that associates a message directly or indirectly with other messages((like hashtags and URLs) or with members of underlying social network	[57]
		Influence factor generated with retweet	[50]
2.	Temporal	Passage of time	[57],[56]
3.	User profile	Complete background of user	[36],[51],[43], [31],[32]
4.	Topical	User's opinion consistency on content related topics	[53]
		User contributed ratings of a product(topic) from consumer comments(user)	[48]
		A particular topic of discussion	[41],[51],[45],[24]

5.	Co-occurrence of words(semantics)	Co-occurrence of words in different contexts	[35],[28],[37], [55],[30]
6.	Local context	The interaction of terms with their neighborhood (modifiers like negations and discourse structure like capitalization)	[34],[42],[58]
		Lexical knowledge and intra-sentential discourse knowledge	[49],[26]
		Correlation between words and sentences	[22]
7.	Global context	Text genre or domain specific context	[34]
		Inter-sentential discourse knowledge	[49]
8.	Textual	Background information related to text (like prior post)	[54],[41], [36],[38],[25]
9.	Dialogical	Sequence of dialogues among a close group of people	[54]
10	Event related	Affective expressions of social media user's under stressful events / public events	[40],[42],[45]
11	NLP features	Linguistic features like n gram, parts of speech tag	[31]
12	Location of post	Location from where post is coming	[43]
13	Domain	Area selected for SA	[29],[41],[44], [58],[31],[27]
14	Appraisal expressions	A basic appraisal unit by which opinion is expressed	[52]
15	Discourse structure information	How the parts of text are assembled	[46]

16	Multiple utterances of video clips	Multiple utterances of video clips	[23]

Number of different definitions were given to a type of context. Different researchers have used different name and definition for the type of context they used in their research. Table 2.4 shows the mapping (context wise) of type of context used and the studies in which it was used.

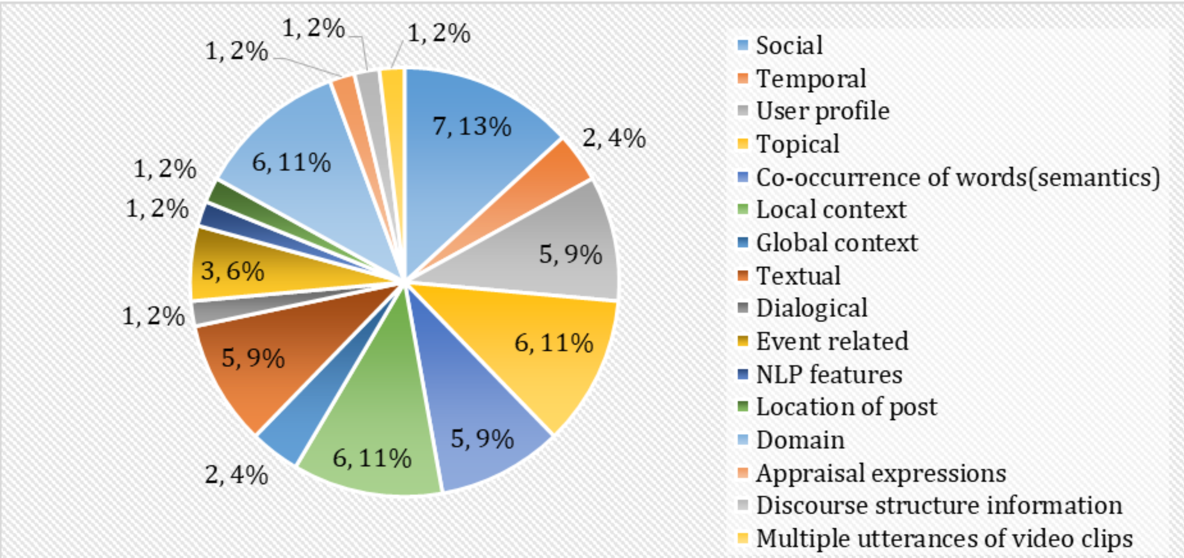


Fig.2.4. Context wise distribution of selected studies

Above figure 2.4 shows the context wise distribution of studies using them. It was seen that social & domain type context were explored more as compared to others. Topical, co-occurrence of words, local and user profile were the next choice of usage to improve upon predictions.

RQ3 intended to find out the techniques used in CBSA. The literature survey in table 2.2 of section 2.3 presents the description of studies which uses context for SA. The following table 2.5 shows the mapping of techniques & studies using them:

Table 2.5. Mapping of techniques with respective studies using them

S No.	Technique used	Study number
1	Lexicon based approach	[42],[30],[37],[55],[35],[34],[43],[27]
2	LR	[39]
3	SVM	[51],[52],[37],[32],[28],[36],[48],[50],[57],[31],[27]
4	NB	[28],[57],[50],[27]
5	Sentic computing	[46],[45]
6	Ontologies	[29],[48]
7	RF	[36]
8	Human Based Qualitative analysis	[54],[40],[58]
9	Dictionary based method	[47],[38]
10	LDA based topic modelling	[48],[24]
11	CNN	[41] [22]
12	CRF model	[49]
13	Maximum entropy	[50]
14	ANN	[50]
15	Fuzzy logic	[44]
16	Quasi Bayesian model	[56]
17	Semi supervised	[33]
18	Matrix factorization	[53]
19	J48 DT	[27]
20	Information retrieval model	[26]
21	LSTM	[22][25]
22	RNN	[23]

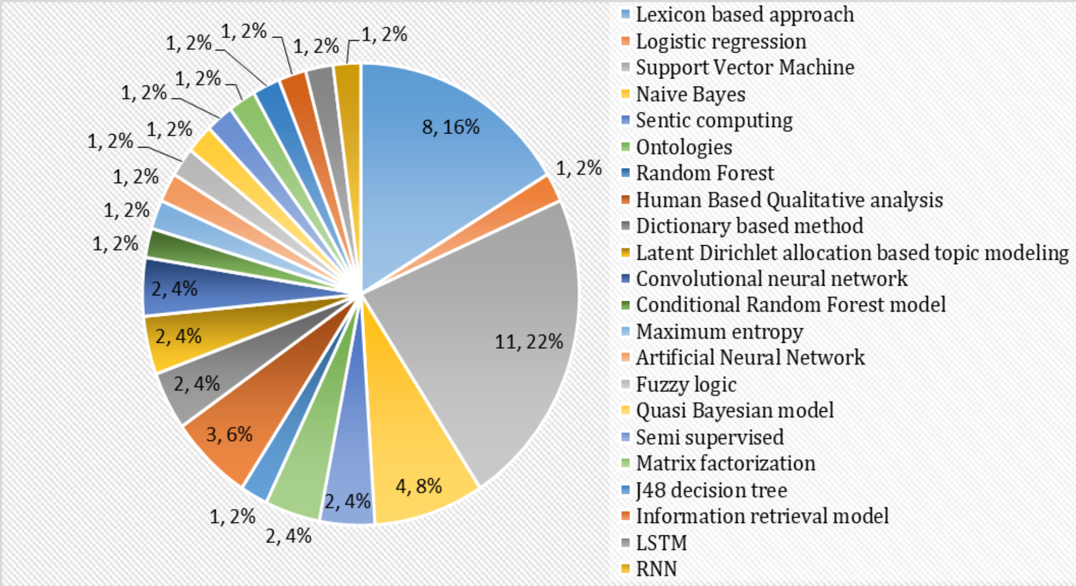


Fig.2.5. Percentage usage of different techniques in CBSA

It was observed that a variety of techniques ranging from lexicon based to ML to the recent deep learning based have been used in CBSA. Figure 2.5 shows that SVM has been a preferred technique with 22% usage in CBSA. Next was the lexicon based approach with 16 % usage. This percentage also included studies using hybrid techniques. On the basis of results presented in table 2.3, it has also been observed that there has been an average improvement of 8-9% in accuracy in results of SA, when context was applied as compared to non -contextual SA.

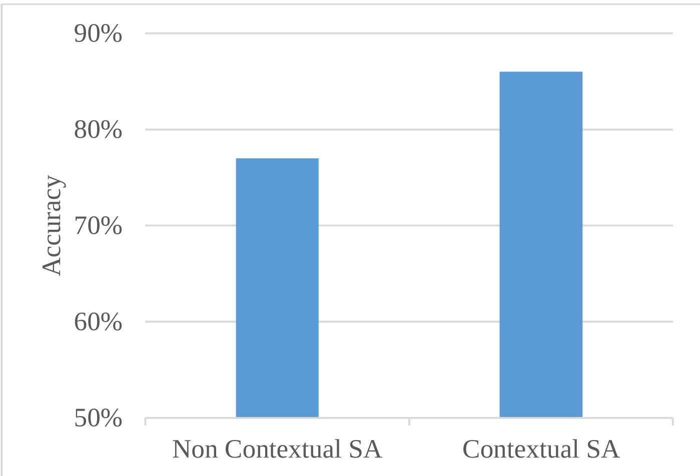


Fig.2.6. Accuracy comparison of contextual and non- contextual SA

RQ4 intends to find the various social media on which CBSA has been applied. Among Facebook, twitter, blogs, customer reviews from Websites, Flickr, Digg, Myspace, IMDb etc. which were used, Twitter was the most used social platform with 38% usage (Figure 2.7).

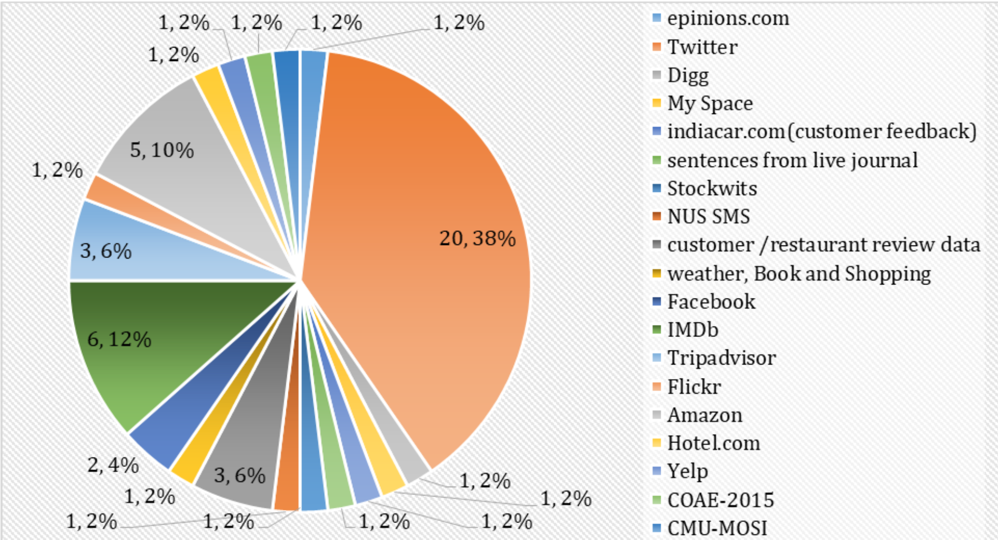


Fig.2.7. Percentage usage of different social media in CBSA

The following table 2.6 shows the mapping of social media used in CBSA and studies using them.

Table 2.6. Mapping of different social media and study using them

S. No.	Social media used	Study no.
1	epinions.com	[48]
2	Twitter	[53],[35],[39],[54],[34],[40],[50],[37],[38],[30],[42],[31],[32],[43],[52],[33],[51],[57] ,[26] ,[24]
3	Digg	[34]
4	My Space	[34]
5	indiacar.com(customer feedback)	[58]
6	sentences from live journal	[37]
7	Stockwits	[28]
8	NUS SMS	[37]

9	customer /restaurant review data	[58],[49],[26]
10	weather, Book and Shopping	[29]
11	Facebook	[47],[36]
12	IMDb	[48],[46],[55],[56],[27],[22]
13	Tripadvisor	[55],[48],[45]
14	Flickr	[41]
15	Amazon	[55],[46],[44],[26],[49]
16	Hotel.com	[27]
17	Yelp(2014,2013)	[22]
18	COAE-2015	[25]
19	CMU-MOSI and IEMOCAP	[23]

Though the SLR was performed for the duration of 2006 to 2018, some pertinent research work was also reported in recent years. In 2019, Fernando et al. [59] gave a formal definition of social context and proposed a framework for classifying and comparing techniques that use social context. In the same year, Zeng et al. [60] proposed a new attentive LSTM model which takes into account the importance of each context word and incorporates the position-aware vectors, which represents the explicit position context between the aspect and its context words. Feng et al. [61] took the microblog conversation as sequence, and developed a context attention based long short-term memory (CA-LSTM) network to incorporate preceding tweets for context-aware sentiment classification. The CA-LSTM network had a hierarchical structure for modelling microblog sequences and allocated the words and tweets with different weights using attention mechanisms. In 2020, Zuo et al. [62] proposed a new context-specific heterogeneous graph convolutional network (CsHGCN) framework that could combine different types of context representations. It had a dependency structure that obtained token-token semantic acquisition more accurately. The proposed model could effectively identify the target sentiment of sentences, and selects qualitatively informative tokens and sentences. Phan et al. [63] too proposed a new approach based on a feature ensemble model related to tweets containing fuzzy sentiment by taking into account elements such as lexical, word-type, semantic, position, and sentiment polarity of words.

2.5 Key observations and research gaps

The SLR enabled uncovering some common/contested observations and important trends in the research area. The following research gaps were identified:

- SA is inherently a restricted NLP problem with following challenges:
    - ❖ Implicit opinions where opinion words are absent or highly ambiguous make SA unintelligible. For example, *"Some people are like clouds and when they disappear it's a beautiful day!"* has an implicit sentiment.
    - ❖ Sarcasm and numerical sarcasm are very frequent in social networks. These are one of the key problems that affect sentiment classifier quality. Sarcasm is defined as a specific type of sentiment where people express their negative feelings using positive or intensified positive words in the text. The idea behind numerical sarcasm is related to changes in numerical values which then affect text polarity. For example, *"It's +25 outside and I am so hot" is* non-sarcastic and *"It's -25 outside and I am so hot"* is sarcastic.
    - ❖ Negation handling is another key problem as it can be implicit, explicit or morphological (prefix: non-, dis-, mis- or suffix-less). It refers to reversing the polarity of words, phrases, and even sentences. It is vital to determine the presence of negation and also the range of the words that are affected by negation words as there is no fixed size for the scope of affected words. For example, in the sentence *"The book was not interesting,"* the scope is only the next word after the negation word. But for sentences like *"I do not call this book a thriller kind,"* the effect of the negation word *"not"* is until the end of the sentence. The original meaning of the words changes if a positive or negative word falls inside the scope of negation—in that case, opposite polarity will be returned.
    - ❖ Word ambiguity is another hitch while working on SA. For example, *"unpredictable"* is a text with neutral polarity. If used in the context of a person's behavior, it carries negative polarity otherwise when related to a

movie plot, it becomes positive. The problem of word ambiguity is the impossibility to define polarity in advance because the polarity for some words is strongly dependent on the sentence context.

- Sentiments change when subjected to different events & stressful situations. At the same time sentiment of a user towards a topic changes over a time so temporal aspect needs to be taken into consideration.
- Researchers have conferred a variety of definitions for a particular type of context which has led to ambiguity in its use. There is an obvious lack of universal definition of context and its types. The multiple facets of context in use needs standardization.
- Multipolarity is an issue where various aspects based opinions are given. For example, *"The camera quality of my new mobile is so cool but the battery life is not too good."* Most of the existing SA models will assign a negative or a neutral polarity to this sentence. To deal with such situations, aspect-based polarity considerations need to be made.
- In CBSA, more work has been done on Twitter as compared to other social media platforms such as Facebook, Flickr and Tumblr.
- The datasets for CBSA are predominantly text based. Multimodal data which includes audio, video, image modalities and their mix have not been reported. Multimodal fusion strategies is another possible direction of future work.
- Though English is the 'lingua franca' of the web, multilingual social media has become exceedingly relevant. People now post and share information in the local language or use it in a mix to English. It is imperative for SA tools to bridge this language divide.

The findings of SLR depicted a shift from traditional SA methods to novel intelligence based methods using context. The existing research gaps in SA and CBSA too foster the need of using context to resolve key computational linguistic challenges. Knowledge is domain-dependent and validity of facts changes as the context switches. Moreover, social media posts are often short, and ambiguous. Topical context or discourse information can resolve the ambiguity. This research thus puts forward the multi-faceted concept of context and its application as a viable solution for enhancing the performance of generic SA task.

2.6 Chapter summary

This chapter presented, a systematic literature review conducted on CBSA to identify research gaps and future scope. A brief description of the key terms such as SA, Contextual information, Subjectivity etc. has been done. Among the several ways to conduct literature review, the SLR following the format given by Ketchenham and Charters was chosen in this research. The review process was divided into six stages viz. formulation of RQs, search strategy, study selection, quality assessment, data extraction and data synthesis. Though, the SLR was conducted for the duration of 2006 to 2018, some of research work that was reported in the area of SA after this period till March 2020 has also been discussed in the chapter.

CHAPTER 3

Multi-faceted concept of context

The contemporary web is about communication, collaboration, participation, and sharing. Currently, the sharing of content on the web ranges from sharing of ideas and information which includes text, photos, videos, audios, and memes to even GIFs. The language and linguistic tone of user-generated content are informal and indistinct. Analyzing explicit and clear sentiment is challenging owing to language constructs which may intensify or flip the polarity within the posts. Our research primarily aims to find out the types of contextual information which can be extracted from social media and can enhance the prediction accuracy of the generic SA. This chapter presents the preliminary work done to understand the what, how and why of using the context in SA. Section 3.1 presents the introduction to the concept of context and CBSA. Section 3.2 puts forward the multi-faceted concept of context. Section 3.3 presents a strength-weakness-opportunity-threat (SWOT) matrix made to demonstrate the effectiveness of CBSA. Chapter winds up with chapter summary in section 3.4.

3.1 Introduction

Undeniably, the cross-platform, cross-lingual, multimodal social web is omnipresent. Popular sites such as Twitter, Facebook, and Instagram have become a critical part of our daily lives to share content, stay connected, and gain insights. The expansive user base on such social networking sites generates voluminous data which can be intelligently filtered and analyzed for building a real-time knowledge discovery framework[5]. The user-generated data can be used by organizations to analyze the trends of market, opinion for the elections, recommendation of the products, and services to the users. SA is one such popular NLP processing task to mine web content. This classification task determines the opinion polarity of the post to comprehend the sentiment and/or emotion specified explicitly. Multiple studies have been conducted to analyze the sentiment of the posts, but the language used by users on the web is a mixture of formal and informal language [64,65]. Detecting the accurate sentiment of the post is not an easy task, and the presence of constructs like sarcasm, irony, and humor makes it exigent even for humans [66]. For example in the tweet, '*Unlike Twitter, LinkedIn is full of positivity. People whom I have never worked with are endorsing me for the skills that I don't possess'* conveys a jest which is difficult to understand without cues. Thus, it is imperative to comprehend supplementary cues from users linguistic input that is aware of 'context' and aids right interpretation.

Context is a set of facts or circumstances that surround a situation or event. Understanding context is one of the most difficult aspects of content moderation. We define context in SA as any complementary source of evidence which can either intensify or flip the polarity of content. Contextual assistance had been studied across pertinent literature, and its effectiveness in SA had been validated. As sentiment reflects more latent information in text, the meanings that sentiment words contain are often context-sensitive. Contextual clues helps to detect fine-grain sentiment from text by resolving the ambiguity of meaning and improve the generic polarity classification of voluminous user-generated social textual data. We defined CBSA as a process of systematic computational analysis of opinions, sentiments and emotions expressed in the textual or non contextual content using contextual clues which can either intensify or flip the polarity of content. The accuracy of polarity classification will thus depend on a context vector, and the learning model will ensure the overall decision making (classification) is more reliable. CBSA is thus a well- recognized task-based solution to improve the conventional SA.

Studies reveal that context is a multifaceted concept with no standard categorizations. Our work presented in this chapter formalizes the concept of context in SA by defining types of contextual cues which may assist fine-grain SA, emotion analysis, sarcasm detection, irony dete+m ction, humor detection, among others. Finally, a Strength-Weakness-Opportunity-Threat (SWOT) matrix is proffered, which determines the effectiveness of CBSA.

3.2 Multifaceted context in sentiment analysis

Contextualization of words is imperative to bridge the gap between what you have experienced and what you are trying to say. Basically, context creates meaning by providing precise and useful information. The polarity shift and other contextual clues can help detect sarcasm, irony, satire, emotion, or humor from text and improve the generic sentiment classification of voluminous user-generated social textual data. We defined the various broad categories of 'context' to complement SA (Figure 3.1).

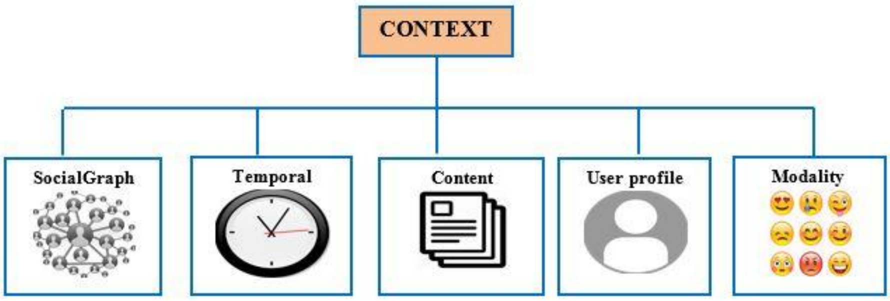


Fig.3.1. Types of *'context'*

These types of context are defined as follows:

- ***Social graph-based:*** Any cue that associates a message or user directly or indirectly with other messages or with members of underlying social network is termed as social context. It can be linked users, connection between messages of the same author, data about likes, Retweet etc.
- ***Temporal:*** Any cue giving time related information is termed as temporal context. It can be origination time of post, passage of time, etc.
- ***Content-based:*** Any cue about the lexical knowledge, i.e. interaction of terms with their neighborhood (modifiers like negations and discourse structure like capitalization), domain, topic, sequence of dialog, and semantics is termed as content-related context.
- ***User profile-based:*** Any cue that is associated with background information of the user is termed as user profile context. It can be personal information, interests, online activity, etc.
- ***Modality-based:*** Any cue from different modalities (text supporting images or images assisting text, emojis assisting text) used in data can be associated with intensification or diminishing the polarity strength.

The SLR conducted within the domain of CBSA further divulged that the maximum amount of research was done using content-based context (textual, global, local, semantic, topical and semantics); whereas, the modality of user-generated content that is visual, typographic, infographic, emojis and acoustic was least explored as context (Figure 3.2).

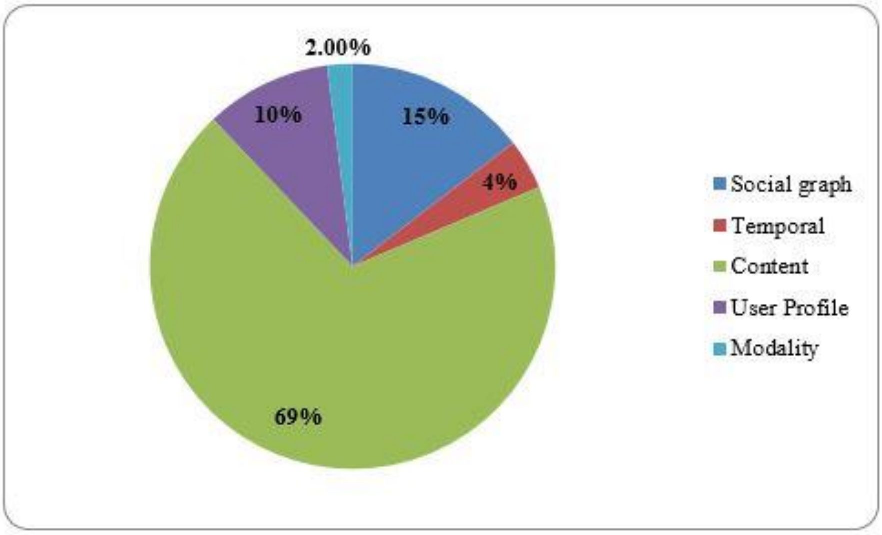


Fig.3.2. Studies conducted on the types of 'context'

3.3 SWOT of context-based sentiment analysis

In this section, we discuss the strengths, weaknesses, opportunities, and threats (SWOT) of using context in SA (Figure 3.3).

3.3.1 Strengths

Context provides additional information about the text which can be used to improve the accuracy of sentiment classification. Various categories of context, viz. language-based, spatial, temporal, social, user profile, etc., can be applied to achieve better results.

- CBSA considers metadata (non-textual features) which is able to contribute greatly to the performance of SA algorithms.
- CBSA is effective both for regular texts and texts with a high degree of noise which do not follow the grammatical rules.
- It is able to detect the shift in polarity of a certain term considering the scenario related to it. Moreover, context gives hint in the background which is able to effectively identify the strength of polarity.
- The concept of context in SA is able to handle issue of polysemy.
- Context is imperative as sentiment polarities are context-dependent. Change in context can lead to variation in polarity.

- Context provides knowledge of domain which is useful in cases when only the text does not provide enough information about the hidden sentiment.
- Shifting of neutral polarity to positive or negative is possible with the addition of context as context is domain-dependent.
- Temporal context if considered is able to identify change in sentiment of a user for a topic with the change in time.
- It enables the integration of information from multiple sources, such as the metadata to identify the polarity of text. The metadata includes identification of influential users, biased users, rumor mongers, the time of creation, the use of emoticons, the length of the text, etc.

3.3.2 Weakness

- Although the concept of context when applied to SA is able to detect the difficult task of sarcasm, it is intricate to deal with language constructs showing orientation toward multiple figures of speech like humor, irony, and satire. These are figures of speech with a thin line of demarcation between them.
    - ❖ Humor is something that is funny and comical. It is aimed at inducing amusement and laughter. Commonly, casual jokes fall in the category of humor.
    - ❖ Irony is aimed at highlighting a contrasting situation or an outcome or behavior that is completely opposite to what is expected.
    - ❖ Satire, as opposed to the other two, is aimed at ridiculing the weakness or shortcomings of someone or someone's work.
    - ❖ For example, a post *'Anushka Sharma was named the hottest Vegan by PETA, while Virat runs a restaurant that serves many non-vegetarian delicacies'* can fall in either of the irony or satire categories.
- Another weakness of CBSA is dealing with constructs which cannot be solved by a single type of context. For example, a post *'summers are so wonderful, the blistering heat, the hot winds and sweaty clothes make a pleasant environment 😖(confounded face)',* cannot be analyzed accurately exclusively on the basis of text. The accompanying emoji also needs to be taken into account.

3.3.3 Opportunities

- CBSA considers context of terms and does not rely on grammatical structures, and hence, it is capable effectively of handling noisy text.
- Slangs, non-standard abbreviations, misspelt words, and colloquial words which are beyond the reach of conventional SA can be handled by the CBSA.
- Addition of context with SA is able to deal with multilingual content.
- CBSA is able to detect sarcasm, irony, and humor into some extent if proper context is known.
- Knowledge of context in SA is able to handle ambiguity in data. Topical context or discourse information has the capability to resolve the ambiguity.
- Many a times the meaning of words varies as per the context in which they are used. This is called polysemy. Polysemy may lead to mis-classification of terms using conventional SA technique. This can be solved by using CBSA.
- SA in automatically transcribed text is challenging due to the fact that spoken language tends to be less structured when compared to written language. Focusing on non-textual aspects of the call such as loudness intonation and rhythm may help in improved sentiment classification.

3.3.4 Threats

Some classic problems of NLP pose a threat to CBSA. These include

- **Co-reference resolution**: It is the task of resolving a mention in a sentence refers to which entity. For example a post *'iPhone's camera quality is much better than Nokia Lumina but it is more reliable'.*
- **Negation handling**: A negative sentiment in a sentence is negated when used in combination with another negative word. Such sentences need to be handled carefully to avoid ending up with sentiments opposite to the writer's intention. For example a post: *'Anita is not a cruel teacher. She did not punish the class for no reason. Students do not have any hard feelings for her'.*
- **Ellipsis resolution**: To make a sentence compact, certain words and phrases are omitted to avoid repetition. For example a post, *'Brothers Billy and John are lucky to have got such intelligent wives who make up for their foolishness. Although Billy recognizes his foolishness, John does not, even though his wife does'.* Such a post

would be more comprehensible even to humans if it was rephrased as Brothers Billy and John are lucky to have got such intelligent wives who make up for their foolishness. Although Billy recognizes Billy's foolishness, John does not realize John's foolishness, even though John's wife realized John's foolishness. Even the context of the sentence does not help clarify who is being referred to here.

- **Slangs and abbreviations**: Social media posts are abundant with slangs and non-conventional abbreviations that are a challenge to any NLP task. For example: *'The weather makes me feel soooohappyyyyy!! I LUUVV it!!'*

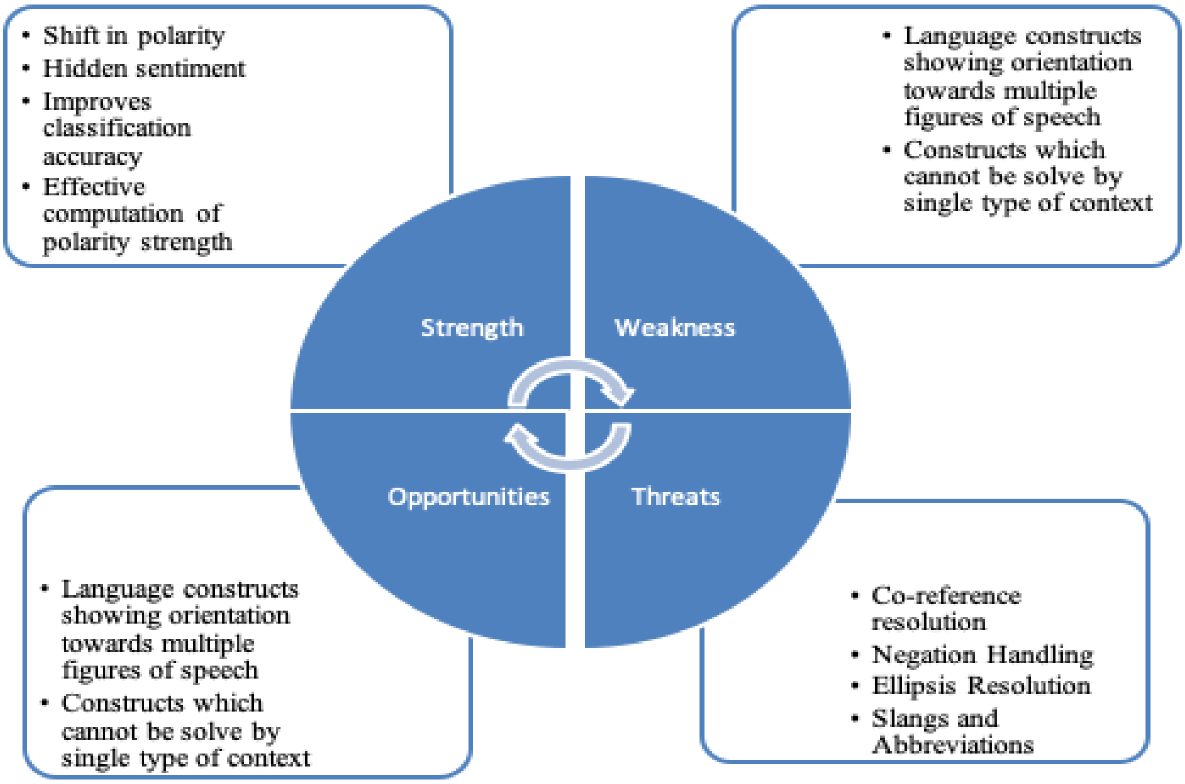


Fig.3.3. SWOT matrix for CBSA

3.4 Chapter summary

Massive amount of data generated by online activity of users can be mined to extract sentiments for usage in different analytical applications. The variety in usage of language construct makes it challenging for automated tools to extract the exact sentiment of the posts. The use of context in SA is one such practical approach which can find use cases to determine sarcasm, irony, and humor in real-time user posts. The chapter discussed the importance of 'context' of the situation, the specific topic, and the environment in analyzing the sentiment by describing the various types of contextual cues which can be put in use, and finally a SWOT matrix demonstrated the research dynamics within the area of CBSA.

CHAPTER 4

Context-based sentiment analysis

Social web users generate a voluminous amount of unstructured data which can be mined to extract sentiments for market, business, and government intelligence. The array of language constructs and usage styles increase the complexity of mining task and call for approaches which can leverage auxiliary add-on information (context). The research work presented in this chapter examined CBSA in both textual and multimodal data. Section 4.1 presents the introduction to the concept. Section 4.2 discusses a context based SA model developed for textual data. Section 4.3 discusses the application of contextual SA model for multiple modality data. Chapter winds up with chapter summary.

4.1 Introduction

As Web 2.0 evolved, websites started allowing user generated content to be created in form of social media dialogues for interaction and collaboration of users in an online community. Social media is inherently an informal way of communication due to the common use of slangs, mal-formed words, and short forms. Mining this massive, high-dimensional, noisy, incomplete social data presents new challenges to all levels of automatic language processing. SA has been thriving to facilitate knowledge extraction for decision making within the omnipresent social web setting. As a generic text classification task, it indispensably relies on the understanding of the human language and emotions expressed via textual or non-textual content. Existing SA techniques quite efficiently capture opinions from text written in syntactically correct and explicit language but shows limited performance while dealing with noisy and implicit content. More recently, as the images, memes and GIFs dominate the social feeds; multimodal content has become a non-trivial element of social media and needs to be analyzed as it has the potential to modify, confirm or grade the polarity of the sentiment. CBSA is the domain of study which deals with comprehending cues which can enhance the prediction accuracy of the generic SA as well as facilitate fine-grain analysis of varied linguistic constructs. We defined CBSA as a process of systematic computational analysis of opinions, sentiments and emotions expressed in the textual or non-textual content using contextual clues which can either intensify or flip the polarity of content. Following sections presents two separate models which were developed in this research, one for dealing with textual data and other for multimodal data.

4.2 Context-based sentiment analysis for textual data

A model using content-based contextual information for carrying SA of textual data was developed in this research. Currently there exists two main types of approach for sentiment detection namely, ML based and lexicon-based. ML approaches use annotated data to train classifiers. Annotation can be performed by asking subject experts to classify each tweet, inferring the emoticons present in a tweet or taking help from sentiment detection websites. The classifiers thus trained become domain dependent and cannot handle new data from a different domain. As twitter contains diverse information from various fields, ML based approaches offer limited applicability.

In contrast, lexicon-based approaches relies on a lexicon to assign a sentiment to the given text and does not require training. Lexicon is basically a collection of words weighted with their sentiment polarities. This approach work fine on texts written in grammatically correct language, but for social media like twitter where content is found to contain misspelled words and slang expressions (for example gud, b8r etc.), traditional lexicons does not suit well. Further, we cannot exploit the lexical structures of the tweet, since the tweets are ungrammatically structured as the message length is restricted to 140 characters only[67].

Our proposed hybrid approach combined methods from ML and lexicon-based approaches for obtaining graded sentiment polarity of the tweets. Generally, tweets are graded as positive, negative or neutral. .We went down one level deeper and classified them into seven classes namely :- highly negative, moderate negative, low negative, neutral, low positive, moderate positive and highly positive. This type of fine-grain grading has been found in emotion analysis but not in SA.

The proposed model works at two levels. At first, a ML based method was used to classify the tweet into one of the three positive, negative or neutral category. The training data required is typically obtained by assuming that positive, negative, neutral tweet polarities can be obtained using emoticons. However, such a training dataset might be incorrect because no standard is defined for the use of emoticons. To resolve this issue, contextual information can be helpful. So, at the second level, a lexicon based approach, which captures context related information, is used for finding sentiment polarity and strength of the tweet. Then polarity and strength of the two levels were combined to assign one of the above seven classes to the given tweet.

4.2.1 Machine learning-based algorithms

This section explains the ML algorithm used in the empirical analysis of the proposed method. Twitter sentiment classification can be treated as a ML task where we have to classify a tweet into three categories, that are, 'positive sentiment' , 'negative sentiment' or 'neutral'. We experimented with the following algorithms: NB classification (as the baseline method) and various tree based algorithms, namely, DT, Random Forest, Extra trees(ET), AdaBoost Classifier, Gradient Boosting(GB) Classifier [68].

4.2.1.1 Naive Bayes

NB is a supervised ML algorithm which applies Bayes theorem[69]. It is assumed that the features taken into account are independent of each other. We take an n-dimensional feature vector x1,..., xn and the probability that it is related to class Y is given as:

$$P(y \mid x_1,\ldots,x_n) = \frac{P(y)P(x_1,\ldots,x_n \mid y)}{P(x_1,\ldots,x_n)} \quad (1)$$

According to the assumption that features x1,...,xn are independent of each other, we can write:

$$P(y \mid x_1,\ldots,x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i \mid y)}{P(x_1,\ldots,x_n)} \quad (2)$$

P (x1, xn) remains constant for given feature vector; we used the following rule to determine the predicted class y^ corresponding to the feature vector.

$$P(y \mid x_1,\ldots,x_n) \propto P(y)\prod_{i=1}^{n} P(x_i \mid y) \quad (3)$$

$$\Rightarrow \hat{y} = \arg\max_{y} P(y)\prod_{i=1}^{n} P(x_i \mid y) \quad (4)$$

Model size in NB is small and generally constant in relation to the data. Due to the small model sizes, NB has lesser computation times. It is good at calculating the multiplication of independent distributions. It is easy to train and work better when categorical variables are less in number. NB is known for its simplicity and generally outperform more sophisticated classification methods.

4.2.1.2 Decision tree

A DT is a model in which the non-terminal nodes represent a particular feature among the set of possible features, arcs coming out of a node are labeled with the possible values of the feature represented by that node, whereas the terminal node state the target class for the input vector that follows the path from entry node to that terminal node. Construction of tree is in recursive divide and conquer manner from top to down [70]. Initially, root contains all the training examples. Categorical attributes are used in the tree (if continuous-valued, they are discretized in advance). Selected attributes guides the partitioning of input data in recursive fashion. Heuristic or statistical measures like information gain guides the selection of test attributes at each node. Classification rules are obtained from the paths in the tree by following root to leaf. An optimal DT is then defined as a tree with minimum height and that covers the most of the data. DT helps in determining worst, best and expected values for different scenarios. It does not suffer from multi collinearity. DT can be easily combined with other decision techniques.

4.2.1.3 Random forest

Random Forests(RF) offer an effective way of classification[71], to overcome the disadvantages of an individual DT, RF constructs multiple DTs which are trained on different subsets of the training data. Each tree is grown as follows

- The number of variable in the classifier be C and the number of training cases be P.
- A number c << C is specified such that at each node, c variables are selected at random out of the C input variables. The best partition on these c is used to partition the node. The values of c is kept constant during the construction of forest.
- Every tree is constructed as big as possible
- The number of votes from all the trees makes the final decision[72,73].

This leads to decrease in over fitting of model on the training data. The outcome of an ensemble model is generally better than the outcome from singleton models. With the growth of forest, it generates an internal unbiased approximation of generalization error. It has an effective method for guessing missing data and maintains accuracy when a large proportion of the data is missing.

4.2.1.4 Extra trees

Geurts et al.[74] proposed another tree-based algorithm which construct an ensemble of DT with extra randomization, hence the name ETs. The training is done like ordinary RF, but the splits or cut-points for a feature are not the local optimal. They are randomly chosen from the range of the values that feature can take. This extra randomization is known to reduce variance to a greater extent. The Extra-Trees partitioning procedure for numerical attributes has two parameters: P, the number of attributes randomly selected at each node and cmin, the minimum sample size for partitioning a node. It is used several times with the complete original learning sample to generate an ensemble model. The predictions of the trees are summed up to produce the final prediction, by majority vote in classification problems and arithmetic average in regression problems[75].

4.2.1.5 AdaBoost

Freund and Schapire [76] used the multiplicative weight-update technique and derived a new boosting algorithm, called AdaBoost. It is basically developed for binary classification. It is treated as one of the most important method in ensemble learning. Ensemble learning involves methods which uses number of learners to solve a given problem.

Weak learners that are models which are a slight better than random chance, such as, DTs, are fitted over the training data and the algorithm assigns weights to such different learners. The weighting mechanism is such that the weak learners try to focus on cases that are more difficult to classify. When the weighted predictions of the corresponding weak learners are combined the overall error is reduced resulting in much accurate final prediction.

4.2.1.6 Gradient boosting

Friedman [77] developed a gradient-descent boosting algorithm. GB consists of following main components:

- A loss function
- A weak learner
- An additive model to add weak learners

Here, after training a weak learner, the loss function for that model is found out. New learner is then fitted over this gradient or the loss function, and the same procedure is repeated with no modification to the previously built learners, unlike AdaBoost where weights of the learners are updated at each iteration. Thus, sequentially an ensemble of tree is formed, where each individual learner is constructed at-a-time, which is then summed to get the final prediction. An advantage of the GB methodology is that an altogether new boosting algorithm does not have to be formed for any loss function that one may want to be used, instead, it is a generalized set up that any differentiable loss function can be used.

4.2.2 Lexicon-based algorithms

This section explains the lexicon based algorithm used in the empirical analysis of the proposed method. In this research, SentiCircles as proposed by Saif et al.[35], a lexicon-based approach, was used in the second level of our SA model. Generally, the lexicon based approaches assign a fixed sentiment score or polarity to the words as given in the sentiment lexicon being used. Such a score is independent of the context in which the word is used in the text. However, SentiCircles assigns polarity and score to the word as given in the sentiment lexicon and also updates the same according to the context. To capture their semantics, it uses co-occurrence patterns of the words in different contexts in tweets.

Traditional lexicon-based approaches suffer from the following main limitations:

- Firstly, they are able to assign polarity and score only to certain fixed number of words which appear in the sentiment lexicon being used. In twitter, new terms, expressions and jargons arise constantly and such a fixed resource may create a problem.

- Secondly, they offer constant and context-independent, word-sentiment polarity and weightage. For example, they would assign the similar score to the word “cheap” in “This mobile phone is quite cheap” and in “He is having cheap behavior”.

SentiCircles are the graph based representations of the words in their contexts [78,79]. Such a representation is dynamic as it updates the pre-assigned sentiment score and polarity from a given sentiment lexicon depending on their contextual semantics. The co-occurrence patterns of words in the text semantics forms the contextual semantics. The textual corpus or the set of tweets is defined as context. The main idea behind the notion of contextual semantics is derived from the fact that a word is known by the environment (textual background) in which it is used. The words co-occurring in a given context have certain semantic influence and relation, which is captured by SentiCircles approach.

4.2.3 Proposed hybrid model for fine-grain grading

Figure 3 shows the proposed model of integrating both ML based and lexicon-based approaches for SA, into a single two-tier frame- work that assigns to each tweet one of the seven classes(high negative, moderate negative, low negative, neutral, low positive, moderate positive and high positive), thus providing a fine-grain grading.

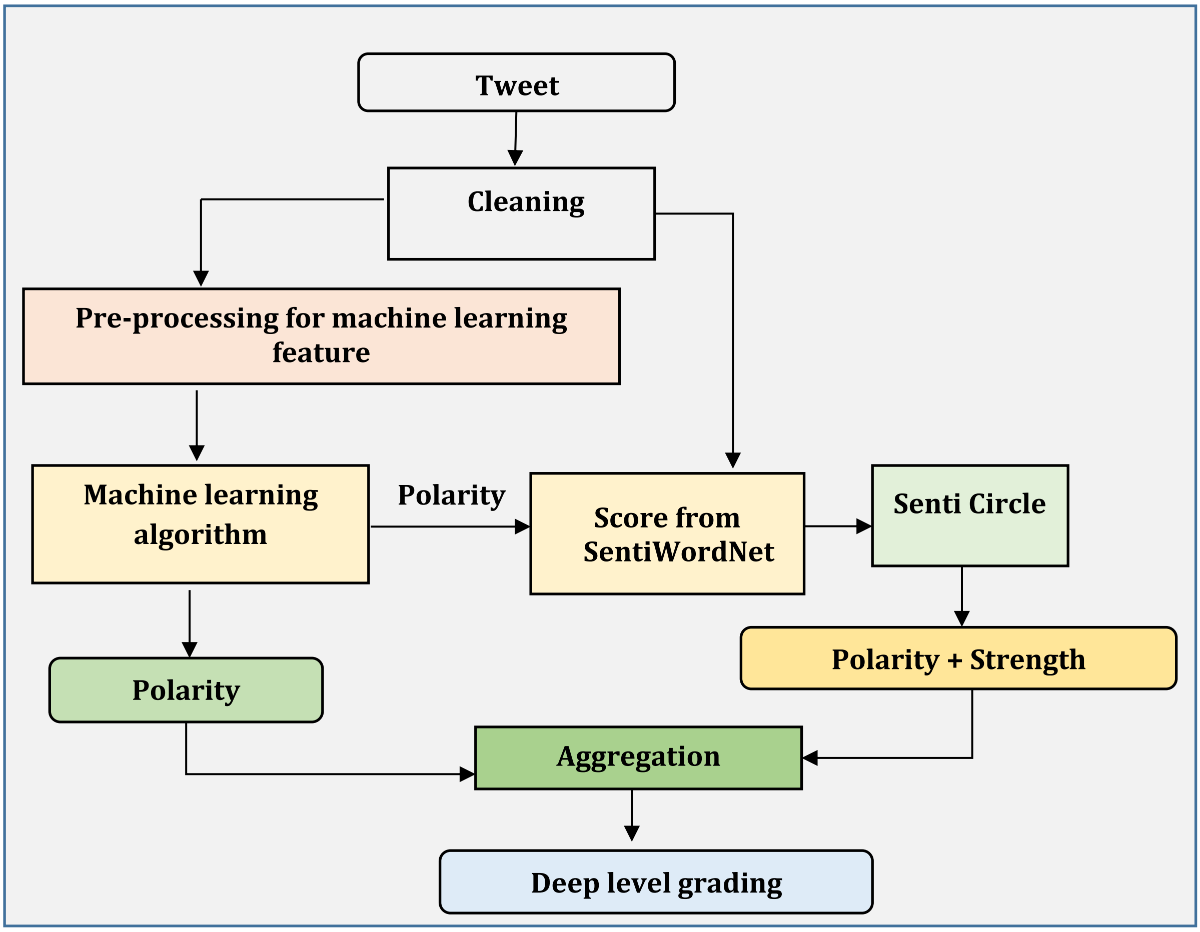


Fig.4.1. Proposed hybrid model

4.2.3.1 Steps in workflow

- **Cleaning tweets:** First we performed cleaning of the raw tweets in the dataset where HTML entities in the tweet were decoded (For example, &amp is changed to &), URLs in the tweet were removed, expressions corresponding to retweet (RT) at the beginning of the tweet were removed, contractions present in the tweet were replaced by their extended words (For example, "I'll" is replaced with 'I will), punctuations present in the tweet including # etc, were removed, three or more repetitive occurrences of a character were replaced with a single character. For instance, 'happppy is changed to 'hapy, terms in the tweet which contains only digits were removed, extra spaces in the tweet were removed, and finally, all the characters of the tweet were changed to lowercase.
- **Feature vector:** Next, to implement the ML algorithms, we used the standard bag-of-features framework. We form a list of tweet words from the tweets in the corpus, which are tagged as a noun, verb, adjective, adverb or pronoun by the Part-

of-speech (POS) tagger provided by NLTK. Now the frequency distribution of each tweet word in this list was found and the top 5000 most common words were taken. These words constitute the bag-of-words which were used as feature words to find the unigrams.

Now, a feature vector corresponding to each tweet was formed. The features used are:

- ❖ Unigrams : presence/absence of feature words
- ❖ Part-of-Speech(POS)features : count of nouns, verbs, adjectives, adverbs, interjections and pronouns
- ❖ Negation : count of occurrences of negation word ‘not’
- ❖ Emoticon features :Various combinations of punctuation marks have been mapped into six classes of emoticons:-Smiley(:),:-), (:), laugh(:D, xD), love(<3,:*), wink(;), ;-D), frown(:-(,:(), and cry(:’(). Count of emoticons belonging to these different classes is taken as feature
- ❖ Count of elongated words(e.g. yummmy)
    - ➢ count of capitalized words
    - ➢ length of message.

Thus, in total there were 5016 features out of which 5000 were feature words, 6 were count of POS tags, 6 were count of emoticon classes, and 6 being other features as mentioned above.

- Ensemble After training the different tree-based classifier listed above, namely, DT, Random Forest, ET, AdaBoost and GB, we took the majority vote among them to determine the final polarity, positive or negative. This ensemble technique [80] allows us to incorporate the advantages of all the individual algorithms, thus achieving improved classification accuracy with increased computation as a trade-off. The evaluated polarity was then used in the second level of our proposed framework as mentioned in the following subsection.
- **SentiCircle** Next we used SentiCircle for Lexicon Based Approach. The steps followed in implementing SentiCircle are as follows:
    - ❖ After cleaning of raw tweet, Firstly, each tweet was tokenized, and each token was POS tagged.
    - ❖ Each token was lemmatized using WordNet Lemmatizer and then stemmed to its root form using PorterStemmer.

- ❖ Based on the POS tag assigned to each token, it was scored using SentiWordNet as follows:
  - ➢ If the POS tag matches one of the tags in SentiWordNet for that term, then all positive and negative scores for that word corresponding to that tag were weighted average inversely according to their sense number separately. Else all positive and negative scores corresponding to that word were averaged.
  - ➢ If positive and negative scores were unequal, then higher of them was re- turned with appropriate sign, else ML output polarity was considered in deciding. If polarity was positive, then positive score was re- turned and if the polarity was negative, then negative score was returned. For neutral output of ML, positive score was returned.
- ❖ Negation Handling: Terms preceded by any of the negative words listed in General Inquirer under the NOTLW category, have the sign of their score reversed.
- ❖ Term-Context Vector Creation: For each word, a vector of words that appear in context of the given word was formed. If the given word appears in any other tweet and that tweet matches the current tweet in having at-least one common user, topic etc., then all the words in that tweet was considered as part of the context-vector of the given word.
- ❖ After forming term-context vector for each term in the tweet corpus, corresponding values of TDOC, θ, x and y were determined for each of the context terms in the context-vector of the term.
- ❖ For each term, its sentiment polarity and strength was calculated by finding geometric median of all its context-terms using weiszfield algorithm.
- ❖ For each tweet, its sentiment polarity and strength was calculated by finding geometric median of all its terms using weiszfield algorithm

- **Combining the results** Once polarity from ML algorithm and polarity and strength (lies between -1 and 1) from SentiCircles were determined, these value were combined according to the figure 4.2 to assign the final deeper level grade to the tweet.

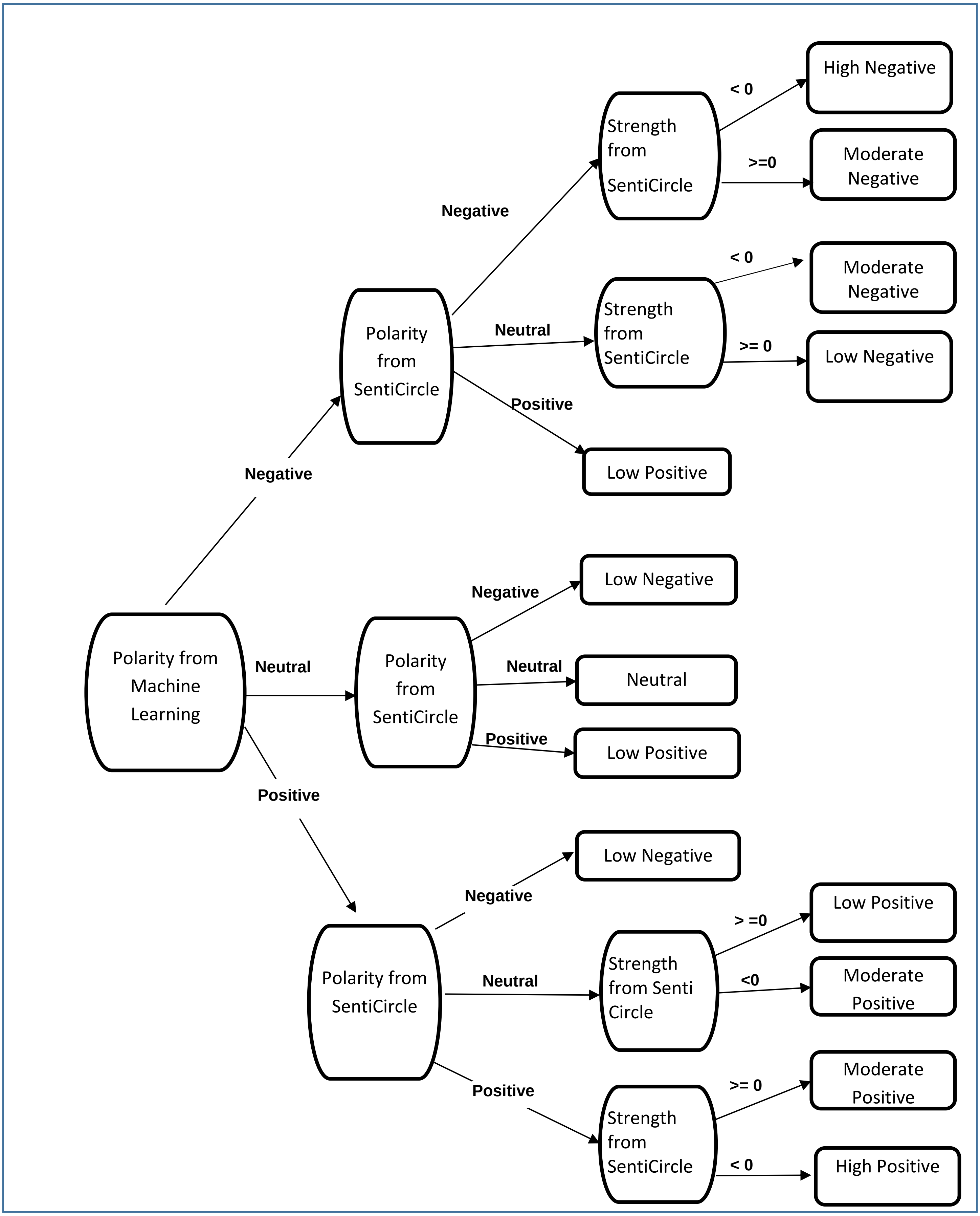


Fig.4.2. Logic for aggregation

4.2.4 Results and analysis

The model for CBSA in textual data used content-based context. That is, textual cues within the given content were used to decipher the intended meaning of use of words.

4.2.4.1 Dataset used

For the evaluation of our model, the STS-Gold benchmark dataset and 3000 tweets from #Demonetization topics were used for the textual CBSA model. STS-Gold is a standard dataset for Twitter SA[35]. This dataset has annotations at tweet as well as at entity level each being independent of the other and therefore supports evaluation of tweet-based and entity based Twitter SA models. It contains a total of 2,206 tweets of which 1,402 are negative, 632 are positive and 77 are neutral.

4.2.4.2 Lexicon used

We have used SentiWordNet[81], a sentiment lexicon, in the second level of our model. In this resource, each word has associated with itself two numerical scores: Pos(s) and Neg(s) which represent the positive and negative score of the entry respectively. Each entry is of the form lemma#pos#sense-number, and the first sense corresponds to the most frequent one. Different word senses can have different polarities. For example, the word cold#a has all the possible combinations, a negative score only, a positive score only and mixed scores as well. Intuitively, mixed scores for the same sense are acceptable, like in cold beer vs. cold pizza.

4.2.4.3 Results on STS Gold & #Demonetization datasets

For STS-Gold dataset, among the tree based classification algorithms, GB gives highest accuracy of 78.01%. All the tree based algorithms have accuracy better than traditional NB algorithm generally used for SA. Lexicon based SentiCircle approach which takes into account the contextual cues, has the accuracy of 55.98% which is further increased to 82.21% by hybridization of ML algorithms and SentiCircle approach for STS-Gold dataset.

Table 4.1. Performance results on STS-Gold dataset

Algorithm	Accuracy
Lexicon Based+ Senticircle	55.98 %
Hybrid (Proposed)	82.21 %
Gaussian Naive Bayes	64.45%
Decision Tree	71.35 %
Random Forest	74 %
Extra Tree	77.39%
Adaboost	77.9 %
Gradient Boost	78.01 %

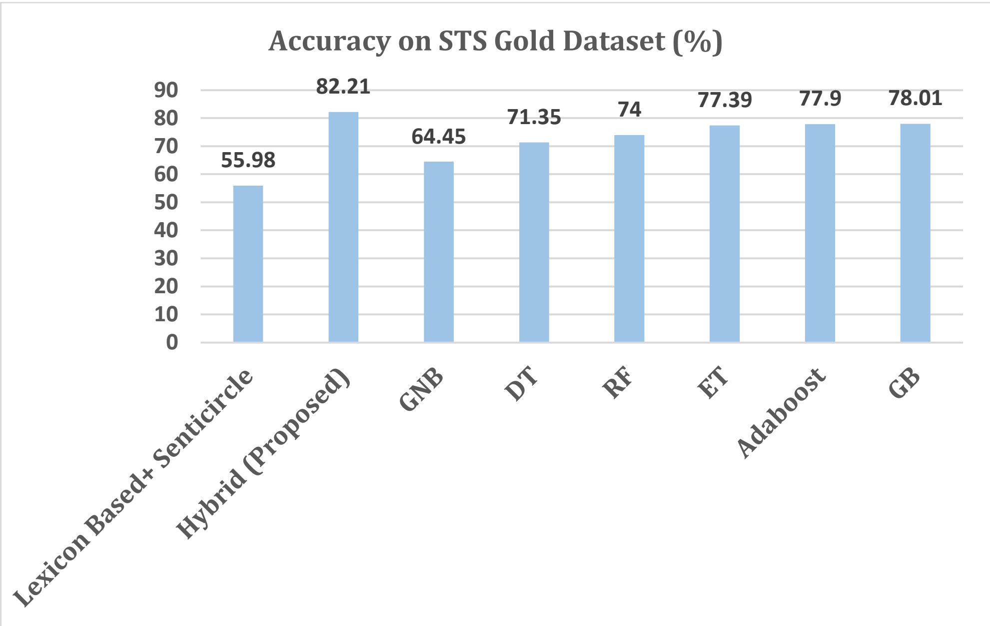


Fig.4.3. Performance results on STS-Gold dataset

Our model showed best performance with accuracy of 87.37 % with topical data (#Demonetization).

Table 4.2. Performance results on #Demonetization dataset

Algorithm	Accuracy
Lexicon Based+ SentiCircle	59. 43%
Hybrid(Proposed)	87.37 %
Gaussian Naive Bayes	68.78%
Decision Tree	74.77 %
Random Forest	76.82%
Extra Tree	78.9%
Adaboost	80.1 %
Gradient Boost	82.53 %

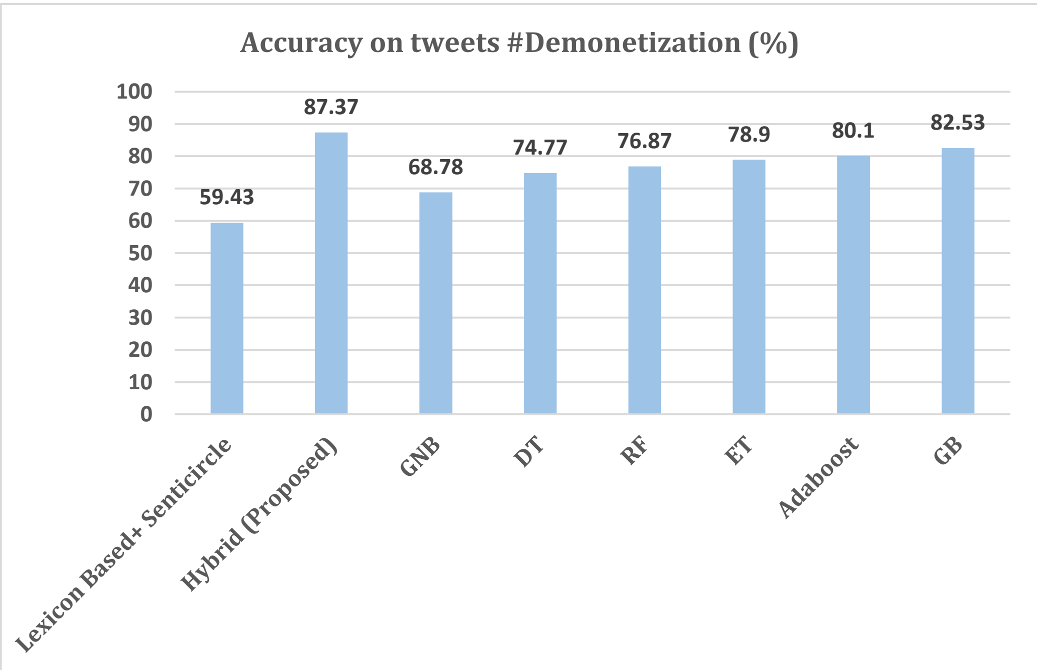


Fig.4.4. Performance results on #Demonetization dataset

4.3 Context-Based sentiment analysis for multimodal data

Text-driven SA has been widely studied in the past decade[12, 14,82,83], on both random and benchmark textual Twitter datasets. Few pertinent studies which report visual SA of images are available in literature [84-90]. But, much of the reported work has analyzed a single modality data whereas multiple modalities of text and image remain unexplored. More recently, as the images, memes and GIFs dominate the social feeds; typographic/infographic visual content has become a non-trivial element of social media. Images are particularly powerful as they have cognition associated and visual experiences convey sentiments and emotions better. Consequently visual SA has been of interest to researchers and it has been observed that deep learning techniques have outperformed the conventional ML techniques in analyzing the visual sentiment. Multimodal capabilities offered by popular social networking websites such as Facebook, Twitter, and Tumblr have further enabled mix of text and images in a variety of ways for better social engagement. The ascendant use of info-graphics, typographic-images, memes and GIFs in social feeds is a testimony to this. Visual content is interesting, engaging and effective. Visual content has both typographic as well infographic content. Typography deals with arranging size, style and weight of the right typoface to provide a visually pleasing format of the text and helps in holding reader's attention. Infography deals with graphic representation of the data to make it easier to perceive.

The multimodal text combines both text and image defining a novel visual language which needs to be analyzed as it has the potential to modify, confirm or grade the polarity of the sentiment. Moreover, human expressions are extremely complicated as statements, images and their mix can convey a wide range of emotions, and often require context to fully understand. Thus, the study to comprehend this text-image relationship is imperative as this combination can modify or enhance the semantics and consequently the sentiment. For example, consider the multimodal text given in the figure 4.5. Here the image of a growling lion depicts a negative, beastly behavior (negative sentiment polarity) which is *modified* by the textual content *"Go Hunt Your dream"* which is a motivational, positive statement (positive sentiment polarity).

Fig.4.5. Example multimodal text with sentiment modification

Now, consider another example of multimodal tweet given in figure 4.6. Here the rainbow colored key- house depicts a happy home with positive vibes and the text *"Love has just checked in"* strengthens this happiness emotion and thus the positive polarity of the sentiment.

Fig.4.6. Example multimodal text with sentiment strengthening

Motivated by this polarity shifts and depths due to content modality, we developed a model to analyze sentiments from multimodal Twitter data which will facilitate visual listening for social media analytics. The model analyzes the incoming tweet for its modality (text, image or image with text) and based on it forwards it to the respective processing module.

4.3.1 The proposed model for multimodal sentiment analysis

The proposed model took into account both the modalities, text and image independently and their combination to analyze the sentiment in tweets. The incoming tweet was firstly examined for its modality type, that is, whether it is an image, a text or a multimodal text (image + text =typographic or info-graphic). The further processing was done on the basis of the identified modality type. For an image only tweet, the image module was implemented which uses an existing model of SentiBank [84] and Regions with Convolution neural network (R-CNN) to determine the sentiment polarity and sentiment score of the image. For a text only tweet, the text module after pre-processing employs a ML based ensemble method (GB) to classify the tweet in to one of the three polarity categories, namely, positive, negative or neutral. Post this; a lexicon based approach (SentiCircles), which captures contextual semantics, was used to determine the sentiment polarity and strength of the tweet. This polarity and strength obtained separately from the ML and lexicon based techniques was combined to do a scoring of sentiment which had the range [-3, 3]. Subsequently, for an image with text, the text was detected and extracted using a computer vision(CV) application programming interface (API) and recognized using optical character recognition (OCR) approach. The recognized text was then processed in the text module whereas the image was sent to the image module for processing. The resultant scores from both these modules were combined to give an aggregate sentiment polarity and score. Figure 4.7 illustrates the systematic flow of the proposed model.

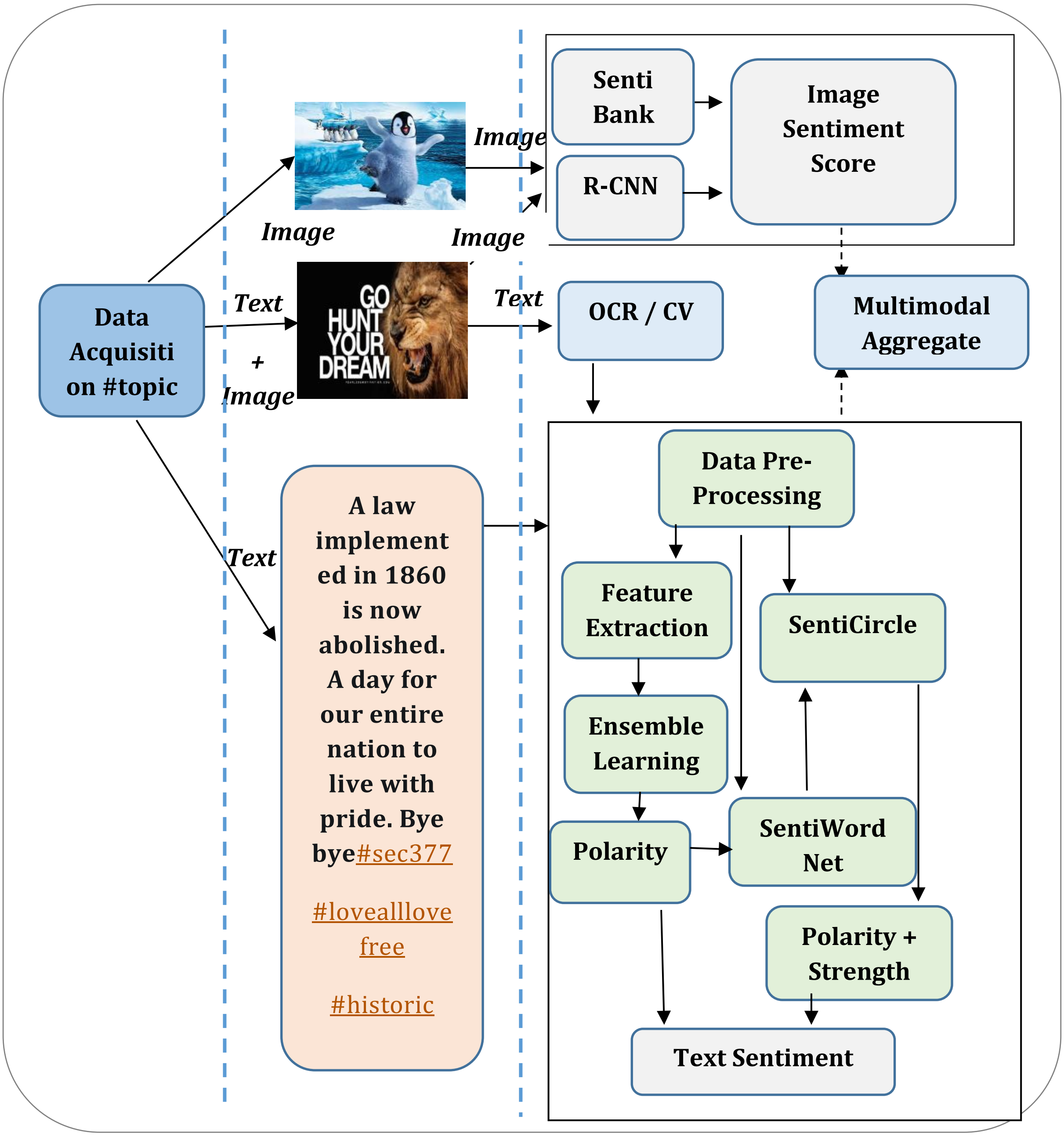


Fig.4.7. Systematic flow of the proposed model

Following sub-sections expound the details of the proposed model:

4.3.1.1 Data acquisition

To evaluate the system using the aforesaid classification techniques tweets pertaining to a topic (#topic) were extracted from the publically available Twitter datasets using its API. 8000 multimodal tweets were collected on the recent topic related LGBT verdict of Indian Penal Court (IPC) section 377 in India using hashtag #section377.

4.3.1.2 Image sentiment analysis

The image only content was processed using the SentiBank, R-CNN and SentiStrength to obtain a sentiment score within the range [-2, 2]. This analytic model to determine sentiment in images was given by Mandhyani et al [91] in the year 2017.

- ***SentiBank:*** It a large-scale visual sentiment ontology which includes 1,200 semantic concepts and corresponding automatic classifiers. Each concept is defined as an Adjective Noun Pair (ANP), where adjective depicts the emotion for a specific object/scene described by a noun [84].
- ***R-CNN:*** R-CNN is one of the popular and efficient object detection models. R-CNN uses selective search for reducing the number of bounding boxes that are fed to the classifier. Selective search uses local cues like texture, intensity, color etc to generate all the possible locations of the object. After this, these boxes are fed to CNN based classifier. In R-CNN the CNN is forced to focus on a single region at a time because that way interference is minimized as it is expected that only a single object of interest will dominate in a given region. The regions are fed to a CNN for object detection.
- ***SentiStrength:*** SentiStrength is a SA program which estimates the strength of positive and negative sentiment in short web texts, even for informal language. Words are classified and rated based on positive and negative strength, i.e., -1 (not negative) to -5 (extremely negative) and 1 (not positive) to 5 (extremely positive).

The pseudo-code used to compute the image sentiment score is given in figure 4.8:

1. Nouns from 1200 ANPs were separated
2. Objects from 200 R-CNN classes were taken
3. The distance between the noun and objects of 200 R-CNN classes was computed by using RBF
4. New ANP weights were multiplied with SentiStrength of adjectives for each ANP.
5. All ANPs were summed-up to get the sentiment score within the range [-2,2]

Fig.4.8. Pseudo-code for image sentiment scoring

4.3.1.3 Textual sentiment analysis

The textual SA is a multi-step process which consists of the following:

- ***Data pre-processing**:* Data pre-processing was done for cleaning and transforming the data for relevant feature extraction. The HTML entities in the tweet were decoded (Eg. &amp is changed to &), URLs were removed, expressions corresponding to retweet (RT) at the beginning of the tweet were removed, contractions present in the tweet were replaced by their extended words (Eg, “I’ll” was replaced with ‘I will), punctuations present including hast-tag ‘#’etc., were removed. Further, three or more repetitive occurrences of a character were replaced with a single character. For example, *‘happppy’* was changed to *‘hapy’*. Terms in the tweet which contains only digits were removed. Extra spaces in the tweet were removed and finally, all the characters of the tweet were changed to lowercase. Additionally, all non-ASCII-English characters were removed, to keep the domain of the data specific to the English language. Part of Speech tagging was also done to extract common structural patterns such as verb, adverb, adjective and noun.
- ***Feature extraction**:* This step identifies the characteristics of the datasets that are specifically useful in detecting sentiments. The classical bag-of-features framework was utilized. We formed a list of tweet words from the tweets in the corpus, which were tagged as a noun, verb, adjective, adverb or pronoun using the Part-of-speech (POS) tagger provided by NLTK [92]. Now the frequency distribution of each tweet word in this list was obtained and the top 5000 most common words were considered. These words constitute the bag-of-words which were used as feature words to find the unigrams. Next, we formed a feature vector corresponding to each tweet. The features used were:
    - ❖ Unigrams: presence/absence of feature words
    - ❖ Part-of-Speech(POS)features: count of nouns, verbs, adjectives, adverbs, interjections and pronouns
    - ❖ Negation: count of occurrences of negation word ‘not’
    - ❖ Count of Emoticon features: Various combinations of punctuation marks have been mapped into six classes of emoticons:-Smiley(:),:-), (:), laugh(:D, xD), love(<3,:*), wink(;), ;-D), frown(:-(,:(), and cry(:’() and their count is taken as feature

- ❖ Count of elongated words(e.g. yummmy)
- ❖ Count of capitalized words
- ❖ Length of message.

- ***Ensemble learning**:* An iterative learning model, GB was then used to train the textual SA module. The GB is meta-model which consists of multiple weak models whose output is added together to get an overall prediction. The evaluated polarity was also fed to SentiWordNet [81].
- ***Lexicon based approach + SentiCircle**:*
  - ❖ Each cleaned tweet was tokenized, and each token was POS tagged using NLTK. Each token was then stemmed to its root form using Porter Stemmer [93]. Based on the POS tag assigned to each token, it is scored using SentiWordNet. SentiWorNet offers a fixed, context-independent, word-sentiment orientations and strengths.
  - ❖ **Scoring from SentiWordNet**
    - ➢ The POS tag matches one of the tags in SentiWordNet for that term, then all positive and negative scores for that word corresponding to that tag were weighted average inversely according to their sense number separately. Else all positive and negative scores corresponding to that word were averaged.
    - ➢ If positive and negative scores are unequal, then higher of them was returned with appropriate sign, else ML output polarity was considered in deciding the polarity. If polarity was positive, then positive score was returned and if the polarity is negative, then negative score was returned. For neutral output of ML, positive score is returned.
  - ❖ **Negation handling** terms that were preceded by any of the negative words listed in General Inquirer under the NOTLW category, have the sign of their score reversed. For example, in the tweet *"Uber Premier is not amazing!"*, the term *"amazing"* is preceded by a negation. Therefore, instead of using its original sentiment score (0.75 in the SentiWordNet lexicon for example, this score negated (-0.75)
  - ❖ **SentiCircle** : SentiCircles considers the contextual co-occurrence patterns to capture conceptual information and update strength and polarity in sentiment lexicons accordingly.

- Term-Context vector was created, that is for each word, and a vector of words that appear in context of the given word was formed. If the given word appears in any other tweet and that tweet matches the current tweet in having at-least one common user, topic etc., then all the words in that tweet were considered as part of the context-vector of the given word.
- After forming term-context vector for each term in the tweet corpus, corresponding values of TDOC, θ, x and y were determined for each of the context terms in the context-vector of the term [78]
- For each term, its sentiment polarity and strength is calculated by finding geometric median of all its context-terms.

- For each tweet, its sentiment polarity and strength is calculated by finding geometric median of all its terms. The polarity from the ensemble learning algorithm and polarity & strength from SentiCircles were then combined to determine the text sentiment score which had the range [-3, 3].

4.3.1.4 Multimodal text: Text in image

For typographic or infographic multimodal text, we used the CV API to extract text using OCR from the image. This process of text retrieval from image comprises of three sub-components viz. text detection, text extraction and text recognition. Text extraction was a crucial step in improving the accuracy and quality of the concluding recognition output. It aims at segmenting text from background that is to isolate text pixels from those of background. An effective text extraction method facilitates the use of commercial OCR without any amendments. For this we have Open CV version**3.4.2,** EAST text detector proposed by Zhou et.al. [94] which is a deep learning model, based on a novel architecture and training pattern. It is capable of running at near real-time at 13 FPS on 720p images and obtains state-of-the-art text detection accuracy. The extracted text was then passed through an OCR for recognition. OCR is conversion of images of typed, handwritten or printed text into machine-encoded text. OCR first pre-process images by techniques like de-skewing, line & word detection, layout analysis, character segmentation etc. to improve recognition accuracy. The character recognition is generally done in two passes. The output of the first pass is transferred to second pass which is a kind of adaptive recognition. Second pass of recognition uses letter shapes recognized with high confidence on the first pass to recognize better the remaining letters on the second pass. OCR sometimes uses a post processing step which makes use of dictionary to improve upon accuracy. Figure 4.9 shows the sample text extraction using the API.

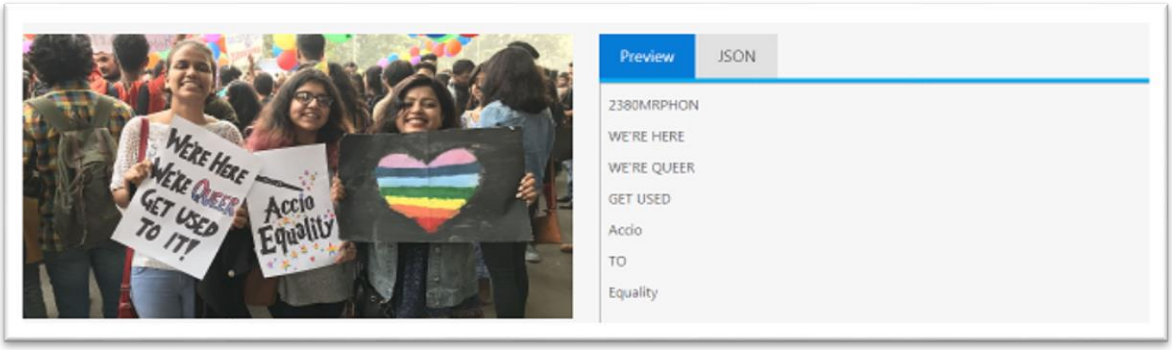


Fig.4.9. Sample text extraction using the CV API

The text thus recognized was then sent to textual SA module for determining the text sentiment score whereas the image was sent to the visual SA module for finding the image sentiment score. These individual scores were combined to produce the aggregate sentiment score for the multimodal tweet. Figure 4.10 depicts the concept of multimodal SA using OCR.

Fig.4.10. OCR and multimodal SA

4.3.2 Results and analysis

To investigate the robustness of the proposed model, the individual text and image modules were validated using benchmark datasets and the model was evaluated for random multimodal tweets. The empirical analysis was thus broadly divided into three parts; (i) Image SA on benchmark Flickr 8k dataset, (ii) Text SA on benchmark STS-Gold dataset, and (iii) Multimodal text (text + image) SA using randomly collected tweets on the selected topic.

4.3.2.1 Result of image sentiment analysis

Image Sentiment was determined using a hybrid of SentiBank and R-CNN. Flickr 8k, a publically available dataset which comprises of images from flicker website was used to train and test the performance of R-CNN for object detection. SentiBank consists of 1200 trained visual concept detectors providing a mid-level representation of sentiment. The results were evaluated initially by only using SentiBank technique and then using a combinational technique of SentiBank with R-CNN. Table 1 depicts the performance accuracy.

Table 4.3. Performance accuracy of image SA techniques

Technique	Accuracy
SentiBank	71%
SentiBank+R-CNN	76.04%

Figure 4.11 shows the results graphically

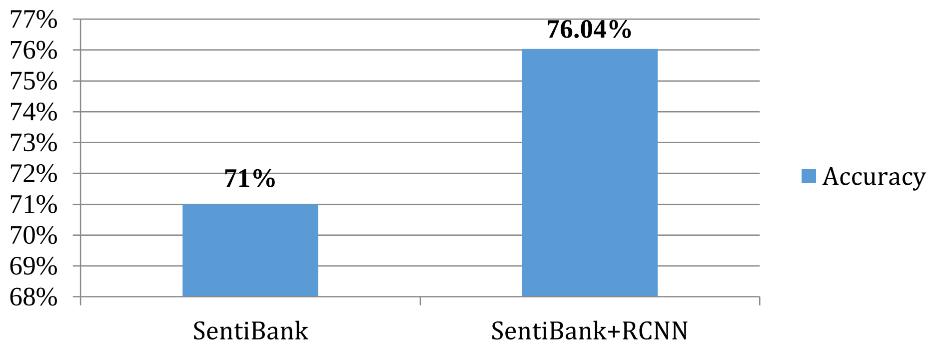


Fig.4.11. Accuracy of image SA techniques

4.3.2.2 Result of text sentiment analysis

For textual analysis step we used STS-Gold, a standard dataset for Twitter SA created by Saif et al. [35]. It contains a total of 2206 tweets, out of which 1402 are negative, 632 are positive and 77are neutral. The performance of the text SA module was evaluated using three approaches, the lexicon-only (SentiWordNet) technique, ML approach (Gaussian Naïve Bayesian, DT, RF and GB) and hybrid approach. Table 2 depicts the accuracy results.

Table 4.4. Performance accuracy of text SA techniques

Approach	Technique	Accuracy
Lexicon Based	SentiWordNet (SWN)	59.43%
Machine Learning Based	Gradient Boosting (GB)	78.01%
	Decision Tree (DT)	71.35
	Random Forest (RF)	74%
	Gaussian Naïve Bayesian (NB)	64.45%
Hybrid	SentiWordNet +Gradient Boosting (SWN+GB)	82.21%

Figure 4.12 shows these results graphically.

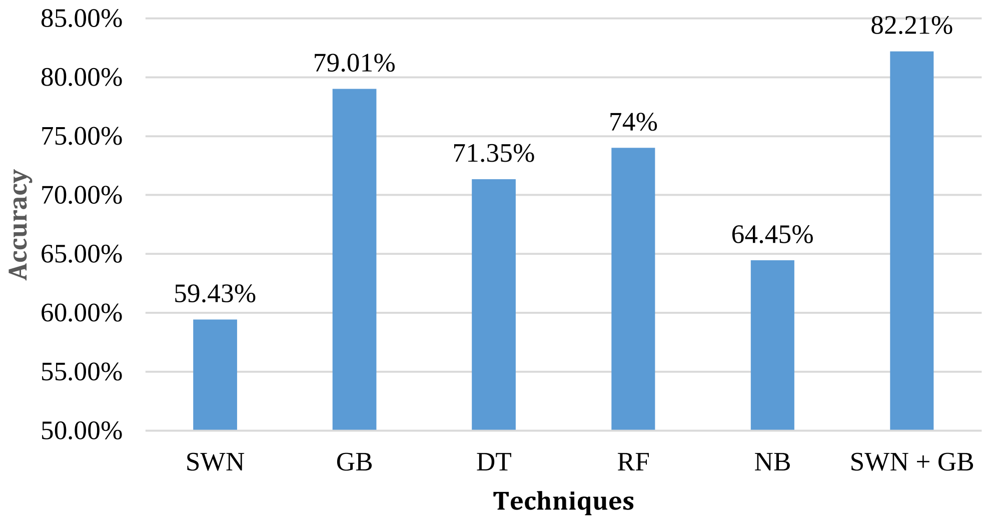


Fig.4.12. Accuracy of text SA techniques

4.3.2.3 Result of multimodal sentiment analysis

Table 4.5 depicts the generic characteristics of text and image tweets:

Table 4.5. Twitter text and image generic characteristics

Text Characteristics	Tweet length	• 280 characters • This limit is Not for Japanese, Korean, Chinese tweets
	Retweet	• No limit on number of times a tweet can be retweeted. • Only recent 100 people who retweeted a tweet will be shown.
	Tweet Limit	1000 tweets/day/Person
	Number of Language supported	40
Image Characteristics	GIF support	• Animated GIFs cannot be included in tweet with multiple images • We can send only 1 GIF per tweet • GIF in full is attached • Photos and GIFs attachment do not count towards character limit • Supports looping GIF
	Image limit	• We can select up to 4 photos to tweet at once • Up to 25 stickers can be attached to a photo
	Format of image tweet	• It accepts GIF,JPEG,PNG • Does not accept Bmp, TIFF
	Size of image tweet	• 5 Mb photo • 5 Mb GIF on mobile • Up to 15 Mb on web

8000 random multimodal tweets on the recent topic related LGBT verdict of Indian Penal Court (IPC) section 377 in India (#section377) were extracted. The distribution of modalities in these tweets is shown in the figure 4.13.

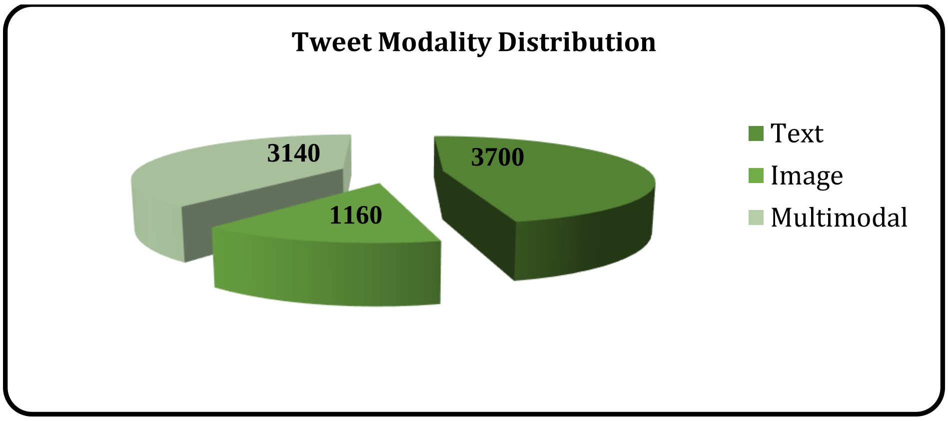


Fig.4.13. Distribution of tweet modality types

The performance of the proposed model was evaluated for these multimodal tweets and the accuracy results for the same are shown in table 4:

Table 4.6. Performance accuracy of proposed model

Module	**Accuracy**
Image module	77.63%
Text module	84.62%
Multimodal(proposed)	91.32%

Figure 4.14 shows these results graphically

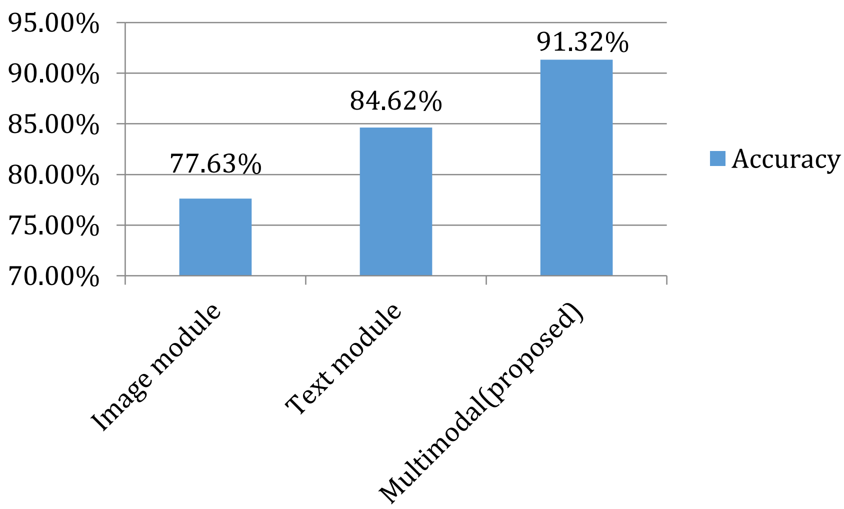


Fig.4.14. Accuracy of proposed model

4.4 Chapter summary

This chapter presents the work done to incorporate contextual information into traditional SA process. A model incorporating contextual information for carrying SA has been discussed in detail. Textual cues within the given content were used to decipher the intended meaning of use of words. An extension of the contextual model to deal with multiple modality data has also been discussed in this chapter. The model for CBSA in multimodal data used modality-based context. The proposed model took into account both the modalities, text and image independently and their combination to analyze the sentiment in tweets. Chapter summary winds up the chapter.

CHAPTER 5

Application of context for implicit sentiment analysis

This chapter presents the research work carried on application of context for implicit SA. It is a preliminary work to understand the what, how and why of using context for SA of one of the type of implicit sentiment i.e. sarcasm used in social media communication. Computational models for sarcasm detection have often relied on the content of utterances in isolation whereas using contextual information definitely improves it. Section 5.1 presents the introduction to the concept. Section 5.2 discusses the Implicit SA: Sarcasm Detection. Section 5.3 presents the literature review on sarcasm. Section 5.4 discusses the Sarcasm Detection using context in benchmark dataset for text only data. Chapter winds up with chapter summary (section 5.5).

5.1 Introduction

The pervasive social web is a rich source of real-time, opinionated user-generated multimodal content available online on resources like review forums and microblogging sites. Automated text analytics using this opinion-rich data has many promising practical applications, such as SA [11,95,96], aggression detection[97], and rumour detection [98]. However, the growing use of emblematic language markers such as punctuations (awesome!!!!!!), emojis (☺, 👎, ❤),wordplay (greattttt for great), creative spellings (2day for today), Internet slangs (OMG for 'Oh My God') [99], code-switching [4] and code-mixing [3] are some commonly seen phenomenon which increase the complexity of computational linguistics to analyze the social media content, thus making SA a non-trivial challenge with a lower prediction accuracy. Human expressions are extremely complicated as statements can convey a wide range of emotions. Figurative languages can easily cheat the SA models with the use of literary devices such as sarcasm, irony and metaphors in the user-generated content. Figure 5.1 represents an instance of a bot misinterpreting a sarcastic tweet and responding inappropriately.

Fig.5.1. Misinterpreted sarcastic tweet

.

Tools to analyze the sentiments get trapped in the intricacies of sarcasm, irony and mixed feelings making it tricky to perceive correct polarity. It requires contextualization of words to bridge the gap between what you've experienced and what you are trying to say. The primary challenges to this SA based social media intelligence include implicit knowledge, nature of text, thwarting and domain specificity, amongst others[100]. Accurately determining polarity from the legit literal text is imperative for real-time text analysis. Tools and techniques to predict accurate sentiment in text are therefore crucial as figurative language permeates daily life and social media[101].

5.2 Implicit sentiment analysis: Sarcasm detection

The subjective nature of opinions makes its mining tricky. Primarily there are four categories of opinions: direct, comparative, explicit and implicit. Out of these direct and explicit opinions are clear and straightforward. Comparative opinion seeks to compare two entities of interest using some criterion. A context analysis is desirable to determine the correct polarity and analyze the sentiment. The stand-alone semantics of the text does not always provide sufficient information about the hidden sentiment that is where there is no explicit opinion word to convey the polarity and the opinions are implicit. For example, the text *"Five minutes of rain means five hours of traffic"* has an implicit sentiment. As one of the most popular literary genre on social media, comedy defines the use of comic literary devices, which include wit, fun, humor, sarcasm, satire, pun, irony and non-sense. We characterized this as the 'Comedy Cube' (Fig.5.2).

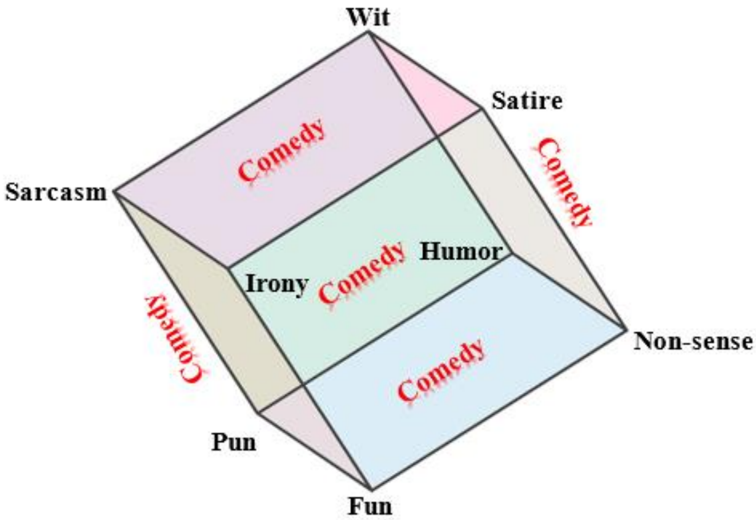


Fig. 5.2. Literary genre of comedy: 'The comedy cube'

Satire is a literary genre that relies heavily on irony, wit, and sometimes sarcasm. *Irony* describes the difference between what is expected and what actually happens and Wit is about understanding and intelligence. *Sarcasm* a sharply ironical taunt; sneering or cutting remark. It is the use of wit or ridicule to taunt, mock or contemptuously strike at a target. Irony is a common tool for sarcasm, but it is by no means the only one. Often, a mere laugh can be sarcastic, without any use of words. With reference to SA, "Sarcasm is defined as a specific type of sentiment where people express their negative feelings using positive or intensified positive words in the text" [102]. It is a manifestation of considerable conflict between the tangible situation and the statement of expression. In a sample post, *"Yummmmmyy!!! Only if you like uncooked and stale food…go and waste your money! 😉"*, this conflict between the actual situation of *"being served uncooked and stale food and is waste of money"* and the expression *"Yummy"* is evident. These inconsistencies within the opinion polarities characterize sarcasm as a special case of SA making the performance of SA task consequential to detection of sarcasm.

Sarcasm detection is a complex task, as it is highly subjective and contextual. Contextualization of words associates experience with expression to add sense. Basically, context creates meaning by providing precise and useful information. The generic content interpretation depends on three non-trivial elements, namely: the text, context and meaning (Fig.5.3) .

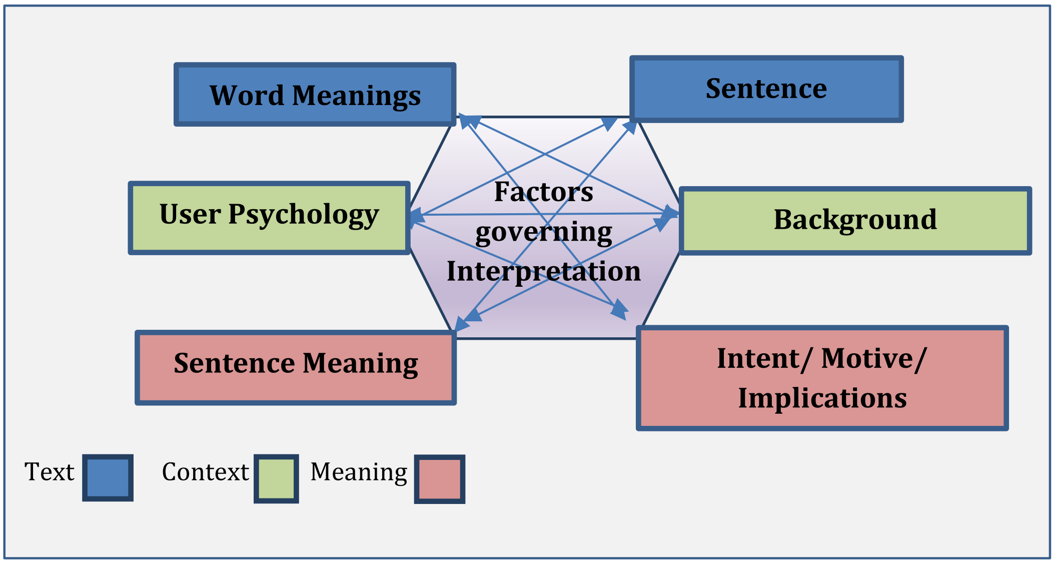


Fig.5.3. Factors governing Interpretation of content

It is imperative to comprehend supplementary cues from users' linguistic input that are aware of 'context' which aid right interpretation. But, elucidating context is one of the most challenging facets in content moderation. Though, contextual assistance has been studied across pertinent literature, its effectiveness in SA to detect sarcasm needs further validation.

Therefore, as a practical use case of CBSA, we probed the problem of detecting sarcasm in social media from the computational linguistic perspective by relying on lexical, pragmatic and semantic cues. The investigation was done considering three potential sub-problems, namely:

- Sarcasm Detection using context in benchmark dataset **(text only data with content-based context to sarcasm)**
- Sarcasm Detection using context for mash-up language **(multilingual text only data with content -based context to sarcasm)**
- Sarcasm Detection in Typo-graphic Memes **(multimodal data with modality-based context to sarcasm)**

The next section provides the related work on sarcasm followed by discussion of sarcasm detection using context for text only data in section 5.4. Sarcasm detection for multilingual text data and multimodal data are presented in subsequent chapters.

5.3 Related work

The contemplation and vocalization of public opinion on online social portals is a goldmine for extracting and analyzing the content available. Twitter has been the quintessential source for mining the sentiment-rich data. 'SA' was first introduced by Dave et al. in 2003 [103] and within the past 15 years the research interest within the domain has increased manifold[5,12]. Literary resources pertaining to practical and theoretical ML based SA on user-generated web content are expansive. Significant studies demonstrate sarcasm in a sentence by conflict and contrast of sentiment polarity[104,105,100]. Detecting sarcastic tone in natural language text is a well-recognized problem within the research area of SA.

In 2010, researchers proposed a model of sarcasm detection in Twitter data and Amazon reviews[106]. The authors used #sarcasm as a filter to obtain sarcastic tweets and used content words and punctuation related features in their work and obtained F-measures of 0.55 and 0.83 for the respective datasets. Another work reported in 2011,[107] divided the twitter data into three types - sarcastic, positive and negative and used lexicon based sentiment scores such as WordNet Affect in their study. In 2013, authors [108] presented an approach for classifying Dutch tweets as sarcastic using unigrams, bigrams and trigrams as features. A model which used a lexicon based approach using the assumption that sarcasm employs a contrast between positive and negative sentiments was also proposed[105]. Another study in 2014, incorporated the intrinsic imbalance between the sarcastic and non-sarcastic texts while modelling their feature set specifically for social media and compared their work with imbalanced classification methods[109]. A hypothesis that Twitter texts are closer to spoken language and hence such features can be incorporated in classification model was also introduced[110] . Ptáček et al. [111] in 2014, worked on a similar problem but emphasized more on Czech language twitter using n-gram and Part-of-Speech features.

In 2015, a Hadoop based framework to capture real time tweets and process it with a set of algorithms which identified sarcastic sentiment efficiently was proposed [112]. In the same year a user-centric approach for detecting sarcasm by designing a behavioral modelling framework fine-tuned for sarcasm was also proposed [113]. Further, a contextual approach to detect sarcasm by including information about the author and the audience and their interaction on the social media platform was given in a study [65]. Researchers in [66] explored intrinsic factors of tweets, including authors' past and use these in a sequential classification approach to detect sarcastic tweets. Another work in 2016 proposed investigated the hashtags #sarcasm, #irony and #not and the similarities and differences between the different figures of language[114]. A two-layered approach to classify French Tweets, by using a generic classifier followed by checking its pragmatic context using search engines to further refine it was also proposed[115].

A number of ML techniques and more recently deep learning techniques have been applied in a variety of research domains[116- 122]. Deep learning techniques attempt to artificially simulate the hierarchical learning approach of the human brain and generalize globally. With the growing popularity of deep learning techniques in NLP task[123], some of these have been exploited within the domain of automatic sarcasm detection too. Felbo et al. in 2017 [124] , proposed a hybrid of attention based Bidirectional Long Short Term Memory(Bi-LSTM) and CNN to detect emotions on Twitter. A variety of features such as emotions; sentiment and personality based were used by authors [125]. Authors in [126]viewed tweets with reference to their contextual environment and use RNN as a part of their model. In another reported work[127], too used similarity between word embeddings as features for sarcasm detection. In the same year, a novel convolution network model was proposed where the model learnt user embeddings and utterance-based embeddings[128]. A combination of CNN, a RNN variant and a deep neural network was given [129] and showed improved performance using deep learning architecture. The study conducted in this research examines ML based and deep-learning based models with a varied set of features for building a robust and efficient automatic sarcasm detection system.

In this chapter , three novel models which were proposed for sarcasm detection are presented. The comparison between the performances of three models in order to find the best model has also been discussed.

5.4 Sarcasm detection using context in benchmark dataset

(**Text only data with content-based context to sarcasm**)

The generic model for predicting a sarcastic text automatically has primarily 4 components, namely (i) Data acquisition and pre-processing (ii) Feature Extraction (iii) Learning using Shallow/Deep Classifiers and (iv)Evaluation using performance measures. The task begins by acquiring the data, which in this work included two datasets, namely the shared task of SemEval 2015- task 11 and Reddit Posts. The unstructured input data has to be converted into an array of representative features to facilitate training and testing of classifiers. Thus, pre-processing task was carried out, which structures the input data for feature extraction. The relevant features (statistical; sentiment-based; punctuation-based; word embedding features) were used to automatically label each tweet/post as sarcastic or non-sarcastic using different classification algorithms. Performance results obtained were then assessed on the basis of evaluation metrics. In this experimental study we investigated three predictive models for automatic sarcasm detection. Each model used, a varied set of features which builds context vector locally and/or globally and was implemented choosing between the set of shallow or deep classifiers. The following sections briefly discuss the models whereas the details of pre-processing, features extraction and classification algorithms used in each model are expounded in the subsequent sections.

5.4.1 First model: Lexical feature + Ensemble voting of shallow classifiers

The first model is based on the Bag-of-Words (BoW) model which uses TF-IDF [8] as a lexical level feature to extract the most descriptive terms in a text. Thus, to assess the significance of a word in the corpus, the TF-IDF features were generated using TF-IDF Vectorizer which was subsequently used as input parameters for training the classification algorithm.

In this first model, three classifiers were used, namely the Multinomial Naïve Bayes, GB and RF. One of the simplest ways of combining the predictions from multiple ML algorithms is by using voting ensembles. A voting classifier was then used to combine model predictions into ensemble predictions. It averages the predictions of the sub-models. This model is the simplest but it does not capture the semantics, position of the terms or their co-occurrences in text. The fig. 5.4 depicts the systematic flow of this model.

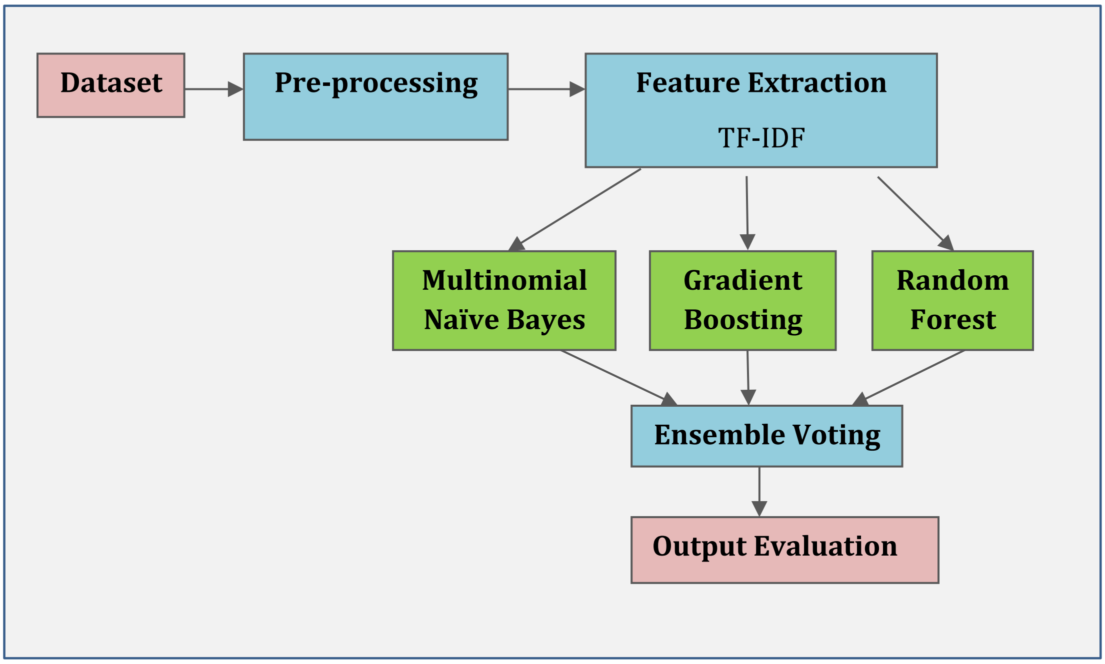


Fig.5.4 Systematic flow of the first model

5.4.2 Second model: Lexical-Pragmatic-Semantic (sentiment) features+ baseline shallow classifiers

The second model combines a variety of features (lexical, pragmatic and semantic). Pragmatic features are linguistic markers that rather than relying on the information a propos the literal meaning of the text comprehend the way of utterance. It describes the "meaning in context" where, context means the situation in which an utterance is made; by whom, to whom, and the immediate circumstances. Prominent categories of pragmatic markers within text include ellipses, quotation marks, exclamation points, question marks, emoticons, caps, parentheses, hyphens and vocalization signals, amongst others. Whereas pragmatics in language is about markers from the point of view of usage, semantics markers are concerned with the study of meaning.

Sarcasm is invariably defined as a specialized sentiment across literature studies. We used the sentiment and punctuation features as defined in the reported work [64] on 'pattern-based approach to sarcasm detection' as semantic and pragmatic features respectively. The details of these features are given in section 3.4.2. These pragmatic and semantic features were then combined with the top 200 TF-IDF features (intrinsic filtering) to generate a feature matrix which was used as input to the baseline classifiers. The fig.5.5 depicts the systematic flow of this model.

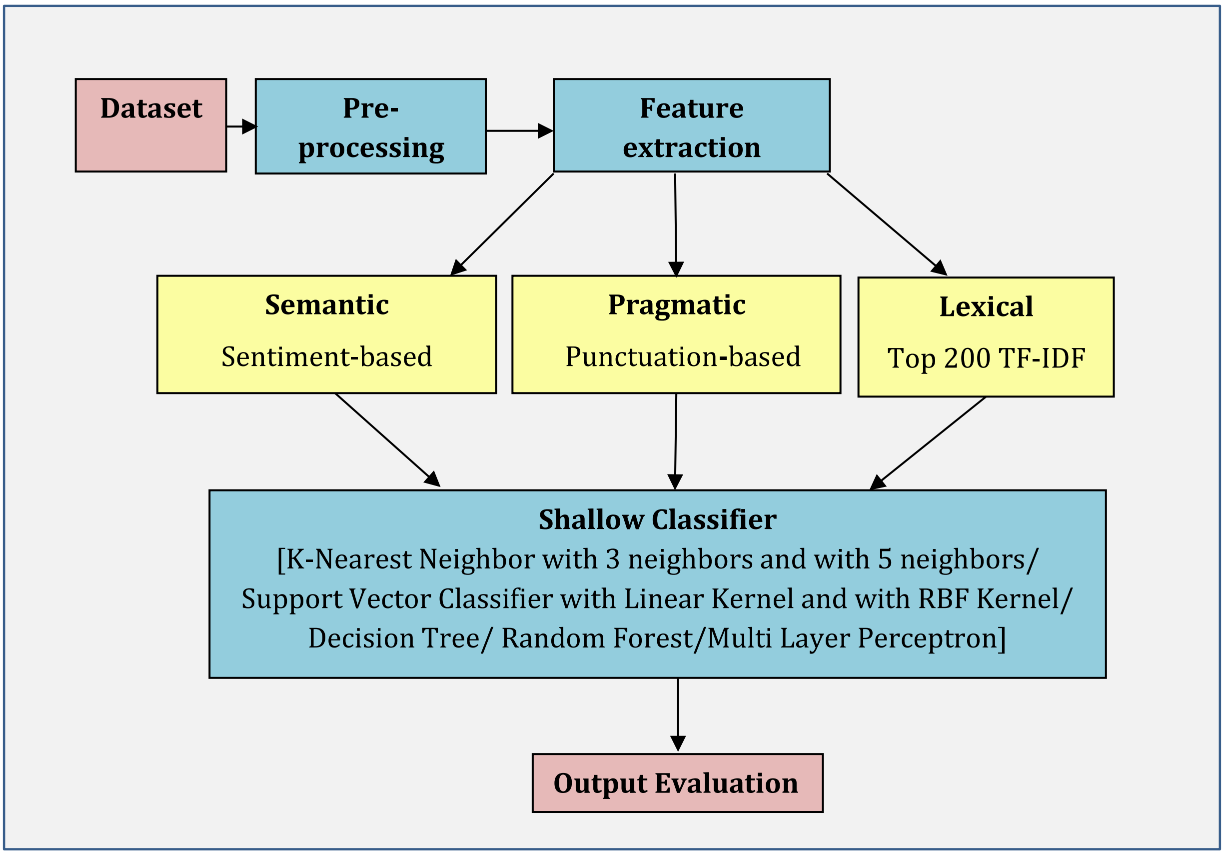


Fig.5.5. Systematic flow of the second model

5.4.3 Third model: Semantic (word embedding) features+ deep learning

Deep learning excels at finding useful representations of the data for a particular task. It has the capability to extrapolate novel features from a limited set of features in a training set, without a human intervention. That is, it searches for and seeks further features that correlate to the previously known ones without the need to label everything [123]. Motivated by the collective intelligence of deep learning, in the third model we demonstrate the use of LSTM and Bi-LSTM models, the variants of RNN on benchmark datasets for improved sarcasm classification. Word embeddings are imperative to any deep learning based NLP task. In this model, the word embedding was done using GloVe[130]. The fig. 5.6 depicts the systematic flow of this model. The dense layer is fully connected layer, so all the neurons in a layer were connected to those in a next layer. The dropout drops connections of neurons from the dense layer to prevent overfitting.

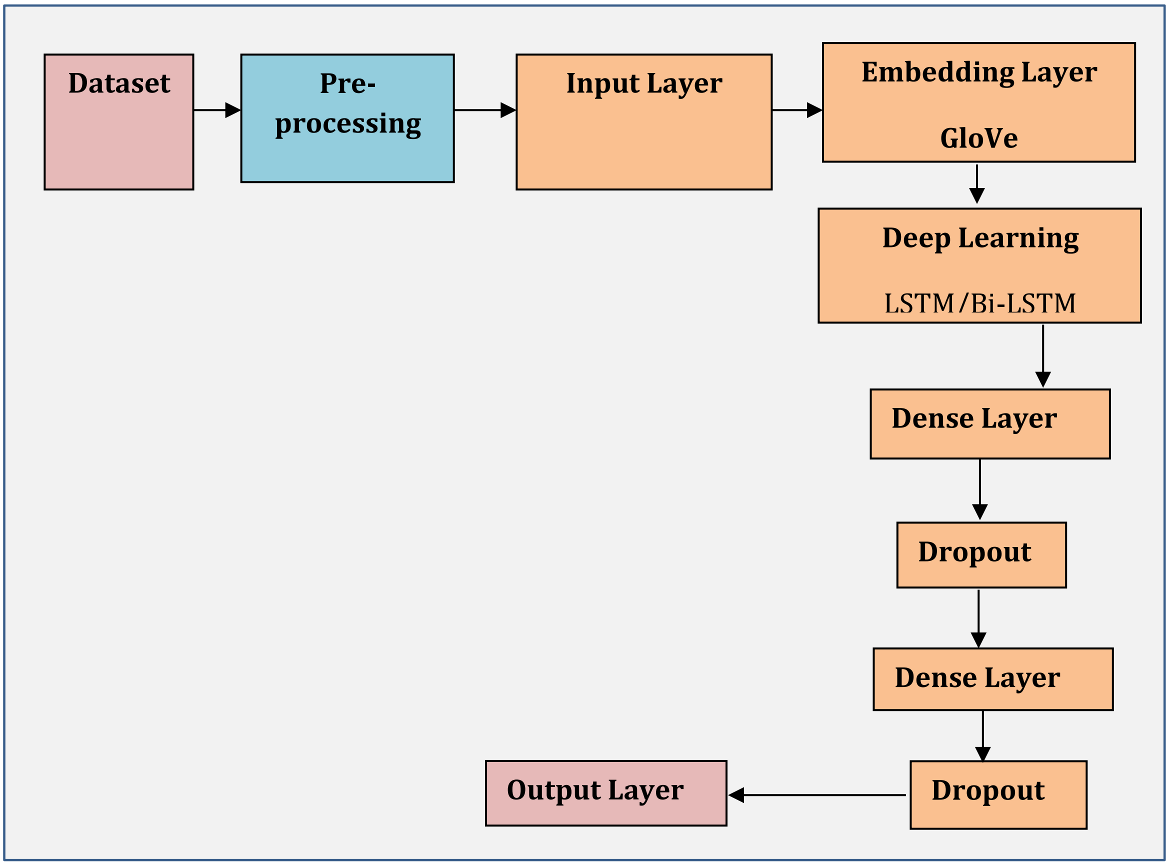


Fig.5.6 Systematic flow of the third model

5.5 Datasets

To evaluate the performance of the three feature-variant predictive models, two datasets were used for empirical analysis. These datasets are, namely, the benchmark Twitter corpus: SemEval 2015 Task 11and the Kaggle's Reddit dataset. The length of the text in the Reddit is greater than the 140 character (now 280) limit set by Twitter, and this was significant for understanding how the models worked on longer texts. This helped in validating the results and in robustness check of the learning models. The table 5.1 presents the various statistics of the two datasets.

Table 5.1. Comparison of statistics pertaining to different datasets used

Dataset	Average length	Average no. of words	Average no. of punctuations	Average length of words
SemEval	104	18	2.4	4.2
Reddit	130	22	3	5

The characteristic features of these datasets are briefly discussed next.

5.5.1 SemEval 2015 task 11

SemEval 2015 task 11 [129] challenge was defined for "SA of figurative language on Twitter". The sarcasm corpus was constructed using hashtags #irony, #sarcasm and #not (considered similar for the empirical analysis). For the negative class data, SemEval 2013 task 2 dataset was used[131]. The dataset is a balanced once as it contains almost equal number of samples from the positive and negative class. A total of 15961 tweets were retrieved, categorized as7994 sarcastic and 7324 non- sarcastic.

5.5.2 Reddit posts

Reddit is a social web portal where users post their thoughts and views on number of topics in pre-defined categories known as subreddits. The "/s" subreddit contains sarcastic content. Hence, posts taken from this subreddit were labeled as sarcastic while others were given negative label. Reddit has no restrictions on the number of characters in a text post and we thus performed our experiments on it to investigate how well our model works on longer text. This was an important aspect since longer text poses the problem of context where sarcasm can only be inferred from a previous statement made in the text. Shorter texts are therefore less "tricky" to classify. The Reddit dataset for this study was taken from the Kaggle competition series [132] and it contains 19737 total posts, categorized as 9999 sarcastic and 9398 non-sarcastic.

5.6 Pre-processing

Post data acquisition, to logically mine the text in tweets & Reddit posts, pre-processing was done to clean and transform the data for extraction of features. This pre-processing included:

- Removal of duplicates tweets, stop words, number in tweets with placeholders, mentions etc.
- Replacement of URLs, special characters such as @, #.
- Natural Language tool-Kit (NLTK) for tokenization.
- Porter's stemmer for stemming to the root word.
- Removal of non-ASCII English character.
- Part-of- Speech tagging was also done to extract common structural patterns such as verb, adverb, adjective and noun.

For example, we consider a sample tweet as follows:

The Pre-processing takes place in the sample tweet as follows:

> "@Raj_232 As a kid, I dreamt of being 30 with a Good CAR, BIG House, and Highly Paid Job. At least one dream came true: I'm 30. https://twitter.com/__yarelii/status/447952784277925888"

- Firstly, the given tweet was broken down into stream of tokens as shown below. These tokens were then sampled in lower case, to maintain uniformity.

> ['@raj_232', 'as', ' a', 'kid', ',', 'i', 'dreamt', 'of', 'being', '30', 'with', 'a', 'good', 'car', 'big', 'house', 'and', 'highly', 'paid', 'job', '.', 'at', 'least', 'one', 'dream', 'came', 'true', ':', 'I', 'am', '30', '.', 'https://twitter.com/__yarelii/status/447952784277925888']

- Next, Hashtags, Numbers in the given tweet, URLs, Mentions were substituted with the placeholders as depicted below.

> ['AT_USER', 'as', ' a', 'kid', ',', 'i', 'dreamt', 'of', 'being', 'NUMBER', 'with', 'a', 'good', 'car', 'big', 'house', 'and', 'highly', 'paid', 'job', '.', 'at', 'least', 'one', 'dream', 'came', 'true', ':', 'I', 'am', 'NUMBER', '.', 'URL']

- Last pre-processing step involved normalization of words by transforming them into singular, active and present tense. This allows easier parsing and accurate feature acquisition.

> ['AT_USER', 'as', ' a', 'kid', ',', 'i', 'dream', 'of', 'being', 'NUMBER', 'with', 'a', 'good', 'car', 'big', 'house', 'and', 'high', 'pay', 'job', '.', 'at', 'least', 'one', 'dream', 'come', 'true', ':', 'I', 'am', 'NUMBER', '.', 'URL']

- Although all the punctuation marks in a tweet were removed as the part of the cleaning process; however, the count of each punctuation mark was kept as we used them as pragmatic features to train the model. In this work, five punctuation-based features that represent figurative text and provide symbolic clues within the tweet were used. These include exclamation marks (!), question marks (?), periods (.), capital letters and use of "or".

5.7 Feature extraction

Three categories of features, namely, lexical, pragmatic and semantic features were used in this work. The first model used a conventional lexical feature, i.e. TF-IDF to train three classifiers that generate a voting based output. The second model used punctuation-based features as pragmatic markers and sentiment-based features as semantic markers.

These pragmatic and semantic features were combined with lexical features to train and test the baseline classifiers. In the final model, semantic relatedness was captured using the pre-trained GloVe word embeddings in deep RNN models, namely the LSTM and Bi-LSTM. A brief discussion on the features used is given next.

5.7.1 Lexical feature: TF-IDF

TF-IDF is a statistical weighting scheme, widely used in text mining tasks. It is a combination of two scores: term frequency and inverse document frequency. Term-frequency is a measure of the frequency of occurrence of a particular term in a single document, normalized by dividing with the total number of terms present in that document i.e. the document length to enable comparisons between documents of different lengths. It is denoted as given in equation 1:

$$TF_{t,d} = f_{t,d} / \sum_{t \in d} t \qquad (1)$$

Where, $TF_{t,d}$ denotes the term frequency of term t in document d,
$f_{t,d}$ denotes the frequency of term t in document d and
$\sum_{t \in d} t$ denotes the total number of terms present in document d

Document frequency, on the other hand, measures the number of documents containing a particular term t. Term-frequency gives equal importance to each term in a document, which can be misleading as certain terms like articles and prepositions occur very frequently in documents but are hardly of any relevance. This shortcoming is overcome by using the inverse document frequency which scales down the weights of frequent terms and gives more importance to not so frequent terms. It is calculated as given in equation 2:

$$IDF_t = \log \frac{N}{\mathrm{dft}} \qquad (2)$$

Where, IDF_t refers to the inverse document frequency of term t
N is the total number of documents in the corpus
df_t is the document frequency of a term t across the corpus

The TF-IDF value was then computed for each term as given in equation 3:

$$\text{TF-IDF} = \text{TF}_{t,d} \times \text{IDF}_t \qquad (3)$$

Thus, the TF-IDF score was lowest for terms that occur frequently in almost all documents, highest for terms that occur frequently in a small number of documents and in between for terms that occur a few times in some documents or occur in many documents.

5.7.2 Pragmatic features: Punctuation-related

Punctuations, wordplay or uppercase alphabets characterize symbolic clues which help comprehend the context within the text. Frequency encoding was done for these pragmatic features as shown in the following table 5.2 [64]:

Table 5.2. Pragmatic features used

rec	frequency of recurring alphabets, that is, if alphabet recurrence> 2, then set the feature to true else false
exclm	frequency of exclamation marks
quest	frequency of question marks
dot	frequency of dots
upper	frequency of capital letters
quotes	frequency of ‘’ or “ ”

5.7.3 Semantic features

While pragmatic features are concerned with the language use, semantic features typically convey the conceptual relationship between words. In this study, two types of semantic features captured the relatedness within words namely, the sentiment and the word embeddings.

- **Sentiment-based features**

  The sarcastic sentences are skewed in the sense that the sentiment polarity of the literal meaning may differ radically from the intended figurative meaning. The shift in sentiment (positive to negative or negative to positive) within these sentences is a strong indicator of sarcasm. For example, in the sentence *“Oh, how I love this feeling of being hurt and alone”* is a clear pointer to a contradicting a fact. In this case, the fact (i.e., *“being hurt and alone”*) and the contradictory statement to that fact (i.e., *"I love"*) express sentiment polarity shifts. Therefore, any type of inconsistency or contradiction between sentiments within the tweet/post can be considered hinting sarcasm. The sentiment-based features defined in a previous reported work[64] were used in this study. The authors characterize two lists of words classified as "positive words" and "negative words" which contain words with the positive emotional content i.e. emotional positive terms (e.g., "love", "enjoy", "happy", etc.) and that with negative emotional content i.e. emotional negative terms (e.g., "hate", "worry", "sad", etc.). These lists of words were created using the pos_tag library under NLTK [92]. Next, using these lists, for each tweet or post the number of positive words (pw) and negative words (nw) were counted. Also, the adjectives, adverbs and verbs have higher emotional content as compared to nouns, therefore, all the words, either positive or negative that have the associated POS tag, were counted another time to create two additional features that represent the number of highly emotional positive terms (PW) and highly emotional negative terms(NW) respectively. Sometimes, emotional content is conveyed through hashtags too. For example, in the tweet, *"Thanks a lot for always helping me #ihateyou",* the hashtag*"#ihateyou"* tells that the user is not actually thanking the addressed user, but was rather extremely disliking him for not helping him. In addition to the already mentioned features, some features that are related to the contrast between the sentimental components were also extracted. Contrast means the coexistence of both a negative and a positive component in the same tweet. The ratio of emotional words was calculated denoted as ρ(t) expressed in the following equation 4:

$$\rho(t) = \frac{(\delta \cdot PW + pw) - (\delta \cdot NW - nw)}{(\delta \cdot PW + pw) + (\delta \cdot NW - nw)} \qquad (4)$$

where,

pw is the count of words with positive sentiment

nw is the count of words with negative sentiment

PW is the count of words with highly positive emotional content

NW is the count of words with highly negative emotional content

ρ is the score determined to find contrast between the above sentimental components t is the tweet

δ is a weight bigger than 1 given to the highly emotional words and is set to 3.

Thus, the sentiment related features considered for this study are summarized in Table 5.3.

Table 5.3. Sentiment-based features used

pw	Count of words with positive sentiment
nw	Count of words with negative sentiment
PW	Count of words with highly emotional positive content
NW	Count of words with highly emotional negative content
ρ score	Finds contrast between the above sentimental components, here delta is equal to 3.

- **Word embedding- GloVe**

  Word embeddings facilitate learned word representations. The benefits of extracting features based on word embedding to detect sarcasm have been recently reported[127]. In our work, for building word embeddings, GloVe [130] was used. It is a count-based model to capture the semantic relatedness of words and builds the learned representation a real-valued vector of words as input. It generates a 'word vector table' representing words by feature vectors. This log-bilinear model studies the relationship of words by counting the number of times they co-occur. Thus, this model aids in mapping all the tokenized words in each tweet/post to its respective 'word vector table'. Proper padding was done for unifying the feature vector matrix.

Thus, the feature set generated in this step was used to train and test the classifier for two pre-defined categories namely, sarcastic and non-sarcastic. A brief discussion on the classifiers used in this study is given next.

5.8 Shallow classifiers

Shallow classifiers refer to the type of task-specific algorithms where the feature extraction is problem specific and the model is trained using these handcrafted features. Shallow structure models generate generalized prediction using supervised learning through classification. Shallow classifiers employ statistical ML models like NB, SVM, DT, RF etc. which perform satisfactorily on the numeric features, but fail to understand the context of the text. Deep learning architectures [such as RNN and CNN] techniques of word embeddings can more efficiently and effectively apprehend the contextual semantics of the information and enhance the generic classification task. The Table 5.4 concisely describes the shallow classifiers used in this study.

Table 5.4. Shallow classifiers used

Technique	Description
Naïve Bayes (NB)	It is based on the Bayes Theorem. Three versions of NB models exists, namely, the Multinomial, Binarized and Bernoulli. In this study, the Multinomial Naïve Bayes explicitly models the word counts and works for discrete data.
Support vector classifier (SVC)	The SVM Classifier (SVC) is described by a decision plane (hyper-plane) which builds decision boundaries for separating group of instances as different class members. The function of kernel is to take data as input and simulate the projection of the initial data in a feature space with higher dimension.
SVC (with linear kernel)	A linear kernel uses linear functions. $$k(x_i, x_j) = (x_i \cdot x_j + 1)^1$$
SVC (with RBF kernel)	It is a general-purpose kernel and is used when there is no prior knowledge about the data. It uses the following equation: $$k(x_i, x_j) = \exp\left(-\frac{\|\|x_i - x_j\|\|^2}{2\sigma^2}\right)$$
K-nearest neighbor (KNN)	K- nearest neighbors is a non-parametric classification algorithm that finds the cluster closest in distance to a particular data point.
Multi Layer perceptron (MLP)	MLP is a type of neural networks which consists of input, hidden and output nodes. It is a self-adaptive and data driven technique.
Decision tree (DT)	It is a ML algorithm that forms a tree based classification where the internal nodes represent features and the leaf nodes represent final classes.
Random forest (RF)	The RF is an ensemble classifier based on Decision Trees. The RF makes a prediction about the class, not simply based on one decision trees, but by an (almost) unanimous prediction, made by '**K**' decision trees.
Ensemble vote classifier	The Ensemble Vote Classifier is used to combine the results of different ML classifiers for classification. It is a meta-classifier and predicts the final class label via majority voting which is the class label that has been predicted most frequently by the classification models.

5.9 Deep learning model

"Deep Learning" also known as "Deep structured Learning" or "Hierarchical Learning" involves learning of deep representation of data be it structured and unstructured and allow to build a solution optimized from algorithm to solve ML problems. Deep Learning has become growing trend to abstract better results when data is large and complex. It consists of a combination of ML and ANN which refers to the depth of the network as it has more than one hidden layer(s).

Deep learning models in literature are defined as neural networks with a large number of parameters and layers in one of four fundamental network architectures:

- Unsupervised Pre-trained Networks
- CNNs
- RNNs
- Recursive Neural Networks

A CNN is basically a standard neural network that has been extended across space using shared weights. A *RNN* is basically a standard neural network that has been extended across time by having edges which feed into the next time step instead of into the next layer in the same time step. It has cycles inside that imply the presence of short memory in the net. A *Recursive Neural Network* is more like a hierarchical network where there is really no time aspect to the input sequence but the input has to be processed hierarchically in a tree fashion. The following fig. 5.7 represents the hierarchy of deep learning and its variants.

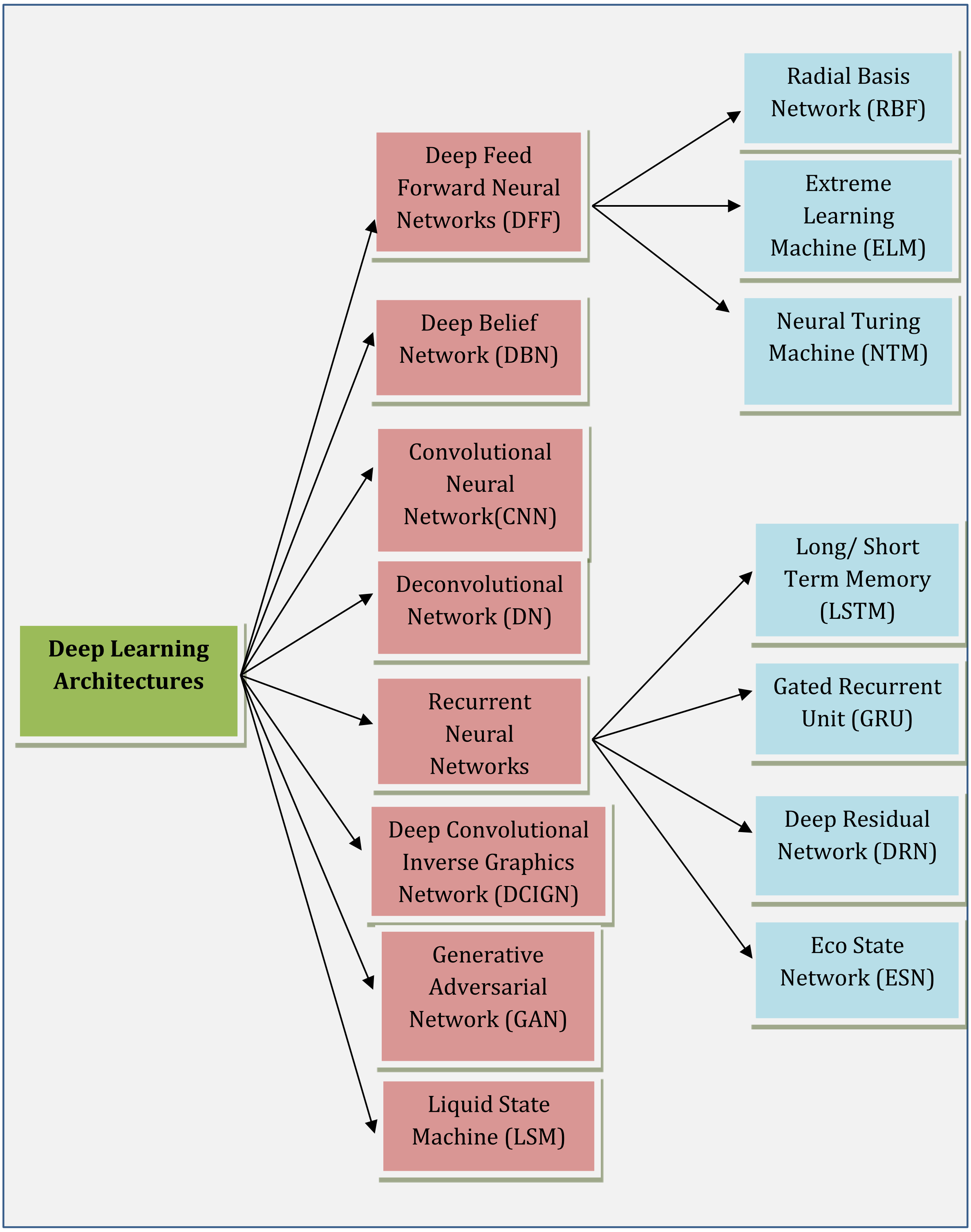


Fig.5.7. Deep learning architectures

As described previously, in this research, LSTM and Bi-LSTM were used to build the deep learning model for sarcasm detection.

5.9.1 Long short term memory model

The LSTM was introduced as a modified version of the classic RNN with the ability to adjust states to determine what to store and what to forget. The RNNs were enhanced with the intent to improve its memory capabilities forming the LSTM [133]. The primary purpose was to deal with the vanishing gradient and exploding gradient problems in the standard RNN. During training, errors back-propagate to adjust weights while maintaining the training properties. Fig. 5.8 illustrate the structure of LSTM.

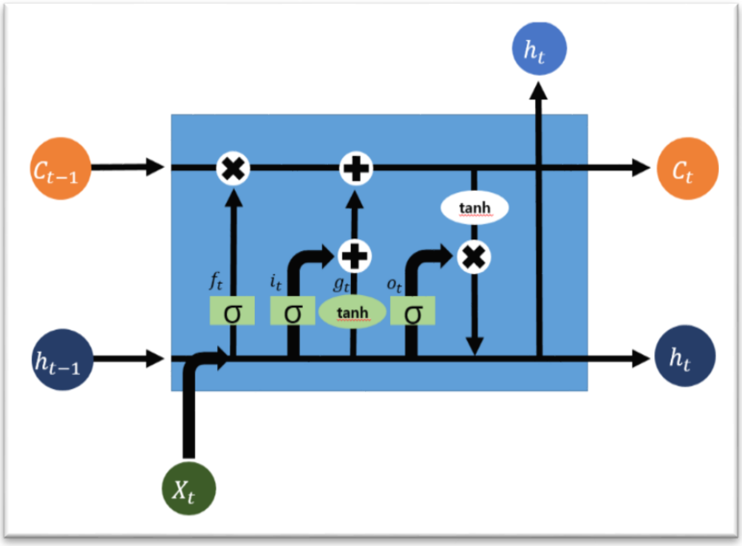


Fig.5.8. Structure of LSTM

The calculation for the forget , the input and the output gate along with the input cell state was done using the following equations 5 to 10:

$$i_t = \sigma\,(w_{i_x} x_t + w_{i_h} h_{t-1} + b_i) \tag{5}$$

$$f_t = \sigma\,(w_{f_x} x_t + w_{f_h} h_{t-1} + b_f) \tag{6}$$

$$o_t = \sigma\,(w_{o_x} x_t + w_{o_h} h_{t-1} + b_o) \tag{7}$$

$$s_t = tanh\,(w_{s_x} x_t + w_{s_h} h_{t-1} + b_s) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot s_t \tag{9}$$

$$h_t = tanh\,(c_t) \odot\; o_t \tag{10}$$

where,

- $\odot$ representselement wise product

- w_i w_f w_o $w_s$are weighing factors used for mapping the hidden layer input to the three gates and the input cell state.
- $b_i b_f b_o b_s$ are bias vectors
- The σ is the gate activation function(sigmoid function)
- tanh is the hyperbolic tangent function.

$Y_t = [h_{t-n}, \dots\dots, h_{t-1}]$ represents the final output of a LSTM layer that is the vector of all outputs.

5.9.2 Bi-LSTM

This consists of two networks where one reads information in forward direction and another reads it in the reverse direction. The output thus generated includes both past and future context. Each element of output vector generated by Bi-LSTM layer was calculated by using the following equation 11[134] (Graves et al. 2013):

$$y_t = \sigma(h^{\rightarrow}t, h^{\leftrightarrow}t) \qquad (11)$$

Where σ function was used to combine the two output sequences. It can be a concatenating function, a summation function, an average function or a multiplication function. The final output of a Bi-LSTM layer was represented by a vector as given in equation 12,

$$Y_t = [y_{t-n}, \dots\dots, y_{t-1}] \qquad (12)$$

Wrapping the LSTM hidden layer with a Bi-directional layer creates a Bi-LSTM. By default, the output values from these LSTMs were concatenated.

5.10 Results

This section presents and discusses the performance results achieved by the classifier using the three models on SemEval 2015 and Reddit datasets based on key performance indicators. Accuracy, Precision, Recall and F1 score were used as the performance measure for each model. "Accuracy is defined as the proximity of a measurement to its true value i.e. it is measured as a proportion of true positives and true negatives among total inspected cases". "Precision defines the exactness of any classifier i.e. it is measured as ratio of true positives to all the predicted positives". "Recall is defined as the ratio of true positives to all the actual positives". "F1 score is the harmonic mean of Precision and Recall"[135]. All the values are expressed in percentages (%).

In order to attain the best performance results optimal selection of parameters is crucial. The Table 5.5 gives the values of parameters used in this research.

Table 5.5 Hyper-parameter values

Hyper-parameter	Value
Dimension of GloVe vectors	200
Hidden units of LSTMs (Forward, Backward)	500 each
Mini Batch Size	10
Regularization	Dropout Operation
Drop-out Rate	0.5: word embedding; 0.2: Bi-LSTM
Learning rate	0.2

5.10.1 Results using the first model

The table 5.6 illustrates the results of the implementing the first model. The results showed an accuracy of 80.23% when the model was applied on the SemEval Dataset while an accuracy of 75.31% was achieved when applied on the Reddit dataset. The fig. 5.9 depicts the results graphically.

Table 5.6. Performance results using First model

	SemEval Dataset	Reddit Dataset
Accuracy	80.23	75.31
Recall	80.74	74.22
Precision	76.39	71.30
F1 Score	78.51	72.73

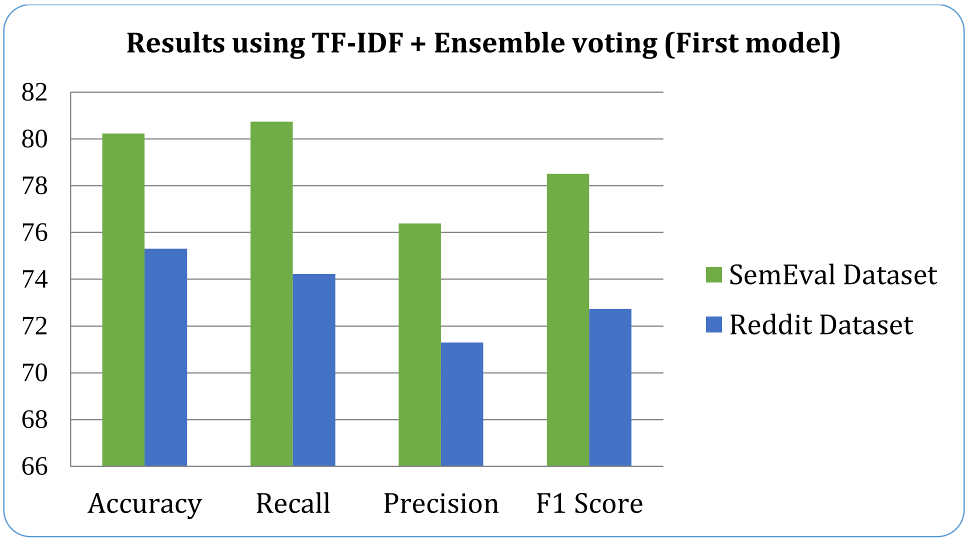


Fig.5.9 Results using the first model

5.10.2 Results using the second model

Our second model extracted 11 features (sentiment +punctuation) which were concatenated with the top TF-IDF features generated. This feature matrix was used to train the five baseline classifier (SVM, KNN, DT, RF and MLP). We considered KNN with 3 and 5 neighbours and the support vector classifier with both Linear and RBF Kernel. The tables 5.7- 5.13 depict the results.

Table 5.7. Performance results using second model (KNN with 3 neighbors)

	SemEval dataset	Reddit dataset
Accuracy	75.61	74.12
Recall	75.83	72.57
Precision	67.11	63.24
F1 Score	71.20	67.58

Table 5.8. Performance results using second model (KNN with 5 neighbors)

	SemEval dataset	Reddit dataset
Accuracy	76.88	77.13
Recall	78.73	79.27
Precision	67.35	69.39
F1 Score	72.60	74.01

Table 5.9. Performance results using second model (SVC with linear kernel)

	SemEval dataset	Reddit dataset
Accuracy	80.72	73.25
Recall	89.56	90.13
Precision	67.43	71.26
F1 Score	76.94	79.59

Table 5.10. Performance Results using second model (SVC with RBF kernel)

	SemEval dataset	Reddit dataset
Accuracy	81.03	75.34
Recall	85.81	77.89
Precision	68.73	64.32
F1 Score	76.32	70.45

Table 5.11. Performance results using second model (DT)

	SemEval dataset	Reddit dataset
Accuracy	73.27	76.37
Recall	67.18	71.38
Precision	67.74	64.59
F1 Score	67.46	67.81

Table 5.12. Performance results using second model (RF)

	SemEval dataset	Reddit dataset
Accuracy	82.69	77.35
Recall	89.13	87.44
Precision	69.08	72.20
F1 Score	77.84	79.09

Table 5.13. Performance results using second model (MLP)

	SemEval dataset	Reddit dataset
Accuracy	83.56	77.95
Recall	80.26	75.39
Precision	72.98	71.83
F1 Score	76.45	73.56

The figure 5.10 depicts the results graphically.

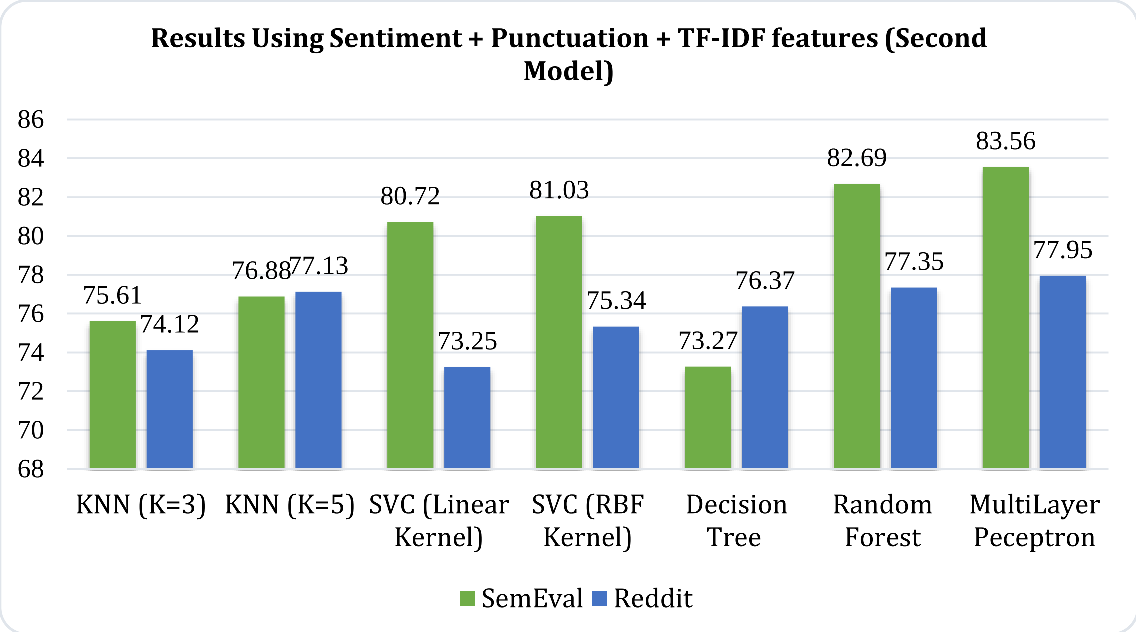


Fig.5.10 Results using the second model

It was observed from the above results that the MLP gave the highest accuracy for the datasets, that is, 83.56 % for the SemEval dataset and 77.95% for the Reddit dataset. Next to MLP was the RF classifier. The least accuracy of 73.27% was observed using DT for the SemEval dataset and 73.25% for Reditt dataset using SVC with RBF kernel.

.

5.10.3 Results using the third model

The accuracy results of this model are represented in Table 5.14 and 5.15.

Table 5.14. Performance results using third model (LSTM)

	SemEval dataset	Random tweets
Accuracy	84.89	80.28
Recall	83.51	78.01
Precision	86.78	74.98
F1 Score	80.39	76.91

Table 5.15. Performance results using third model (Bi-LSTM)

	SemEval dataset	Random tweets
Accuracy	86.32	82.91
Recall	84.39	80.27
Precision	81.61	78.76
F1 Score	85.25	79.45

It was observed that the Bi-LSTM outperformed the LSTM when implemented on both the SemEval and the Reddit datasets. In case of SemEval dataset, it gave accuracy of 86.32% while in the case of Reddit dataset it gave the accuracy of 82.91%. The graphical representation of results is given in the figure 5.11.

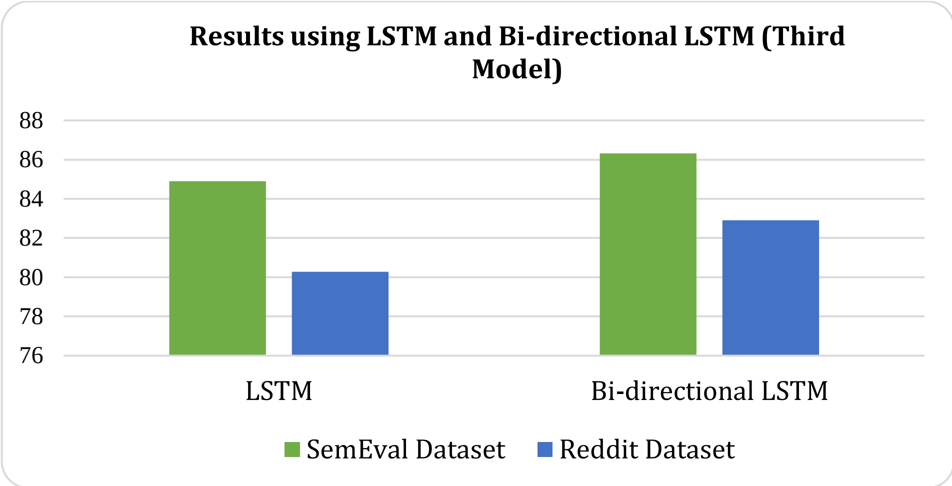


Fig.5.11. Results using the third model

To better understand the deep learning architectures for sarcasm detection, the results of LSTM and Bi-LSTM were compared with CNN too. The classification accuracy was observed as: CNN<LSTM<Bi-LSTM as shown in the following fig. 5.12.

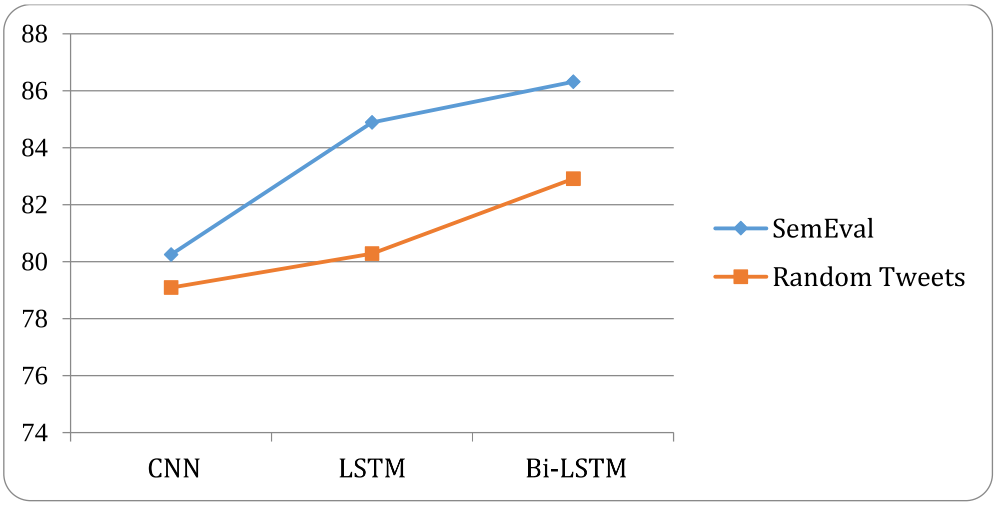


Fig.5.12. Accuracy of deep models

5.10.4 Comparison of the three models

The Bi-LSTM achieved the highest classification accuracy of 86.32% for the Twitter dataset and 82.91% for the Reddit dataset to detect sarcasm. The second-best results were obtained using LSTM with the accuracy of 84.89% for Twitter Dataset and 80.28% for Reddit dataset. The fig. 5.13 depicts the comparison graphically.

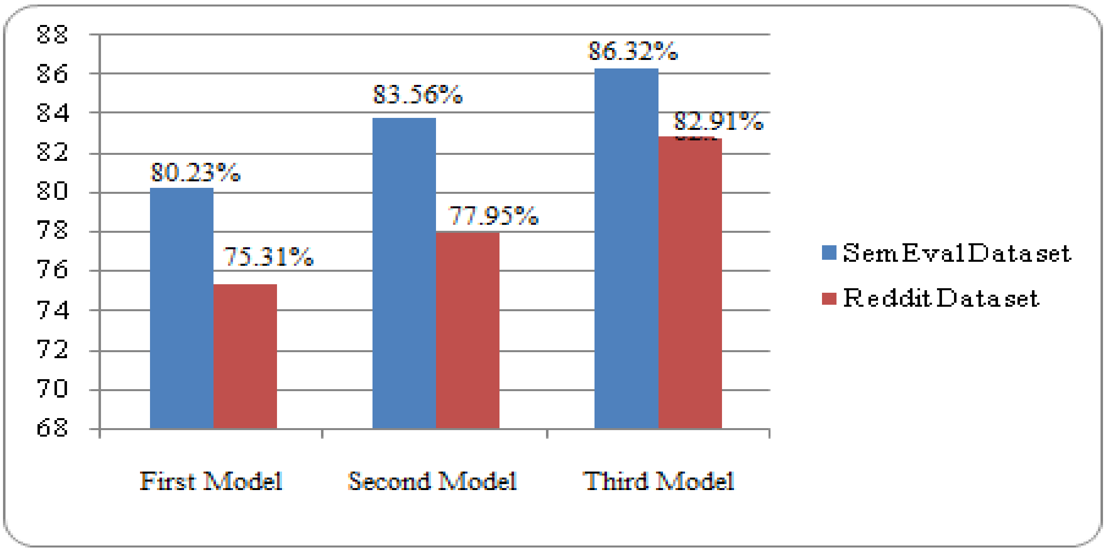


Fig.5.13. Comparison of the three models

5.11 Chapter summary

Automatic sarcasm detection as a typical text classification task has the primary goal to distinguish between sarcastic or non-sarcastic textual content. SA is certainly one of the areas that can benefit most from sarcasm detection. The work presented in this chapter probed the problem of detecting sarcasm in social media from the computational linguistic perspective by relying on lexical, pragmatic and semantic cues. Context was defined using various features such as n-grams with TF-IDF encoding, frequency encoding for punctuations, hand-crafted features and pre-trained word embeddings. Three models were proposed using these linguistic context markers to train and test shallow as well as deep classifiers which predict sarcasm effectively and efficiently in two experimental datasets. It was observed that the Bi-LSTM using pre-trained vectors from the GloVe word embedding is the best sarcasm prediction model with an average accuracy of approximately 85% (~84.6). The preliminary results are clearly motivating. Deep learning architectures have been used for various NLP tasks and can be applied to random, live twitter streams for enhanced classification performance.

CHAPTER 6

Application of context for sarcasm detection in multilingual data

Analyzing explicit and clear sentiment is challenging owing to the growing use of emblematic and multilingual language constructs. This chapter presents the research work carried on application of context for the detecting sarcasm in multi-lingual data. Section 6.1 presents the introduction to multilinguality prevalent in social media posts. Section 6.2 presents the literature review on multilingual sarcasm detection. Section 6.3 describes data acquisition and preprocessing of mash-up Hinglish(multilingual) data. Section 6.4 describes the feature extraction for training the model. Section 6.5 demonstrates the architecture and working of the proposed $^{\text{softAtt}}$Bi-LSTM-$^{\text{feature-rich}}$CNN model for sarcasm detection using context for multilingual text only data. Section 6.6 presents the result analysis of proposed model follwed by chapter summary in section 6.7.

6.1 Introduction to multilinguality

The number of social network users worldwide continue to grow and is expected to pass 3 billion in 2020. The social, interactive computer-mediated technologies, such as, Twitter, Tumblr, Google+, Facebook, Instagram, Snapchat, YouTube etc. which enable users to create, post and share all types of multimedia text are gaining users globally. Social media serves as a communication channel and a social listening, awareness, activism & feedback tool for stakeholder engagement and cooperation. Keeping tabs on social media sentiment, that is, *'the online mood'* is a key part of social media listening. Determining the literal exactitude and opinion polarity in text is imperative for real-time SA. The non-standard vocabulary with informal textual content and noise currently defining the content strategy of social media[136,97,98] makes automated text analytics computational intensive.

Multilinguality is a well-recognized challenge of NLP. The cultural diversities, country-specific trending topics and hash-tags on social media and easy availability of native language keyboards for social media applications add to the variety and volume of user-generated content in diverse languages and dialects [137].The mixture of languages can be observed in text as follows:

- *Transliterated Code-mix*: I looooovvveed the movie!!! Dil aa gaya…ek dum super ❤
- *Literal Code-switch*: I looooovvveed the movie!!! दिल आ गया …एक दम super ❤

The first case is an example of transliterated bilingual code-mixing where one language/script word (Hindi) is transcribed into a source language (English) such that the source phonetics is preserved. This is also known as phonetic typing. The second example describes a literal bilingual code-switch, where the actual words of one language (Hindi) are mashed up with the other language (English) demonstrating language alternation. Much of the recent research has been conducted on transliterated code-mix using publicly available API or by building subjective lexicons [138]. SA research too has either been conducted on individual languages [139] or transliterated code-mix [140]. Also, specialized sentiments like emotion and sarcasm detection in tweets has been primarily done on monolingual English tweets where the non-ASCII words are removed during the data pre-processing step.

Recent statistics, show that Hindi is the fourth most widely spoken language around the world with about 310 million native speakers, coming in only after Chinese, Spanish and English. An upsurge in the use of hybrid of Hindi and English languages has been observed [141]. The availability of keyboards with 'Devanagari' scripts on mobile phones has made it a popular language choice. A research study from 2016 by Parshad et al. [142] shows that people are more fluent in Hinglish than in English alone. The research presented in this chapter, focuses on sarcasm detection in the mash-up language (literal code-switch) tweets, specifically the Hinglish, which refers to the juxtaposition of words from Hindi and English language. For example, "*Well played AUSSIES!!!* एक और हार LOL 👏👎". The shift in polarities or intensification was analyzed for detecting sarcastic tone in Hinglish mash-up tweets.

As a typical natural language text classification task, automatic detection of sarcastic tone depends on feature engineering and learning model. The hierarchical learning capabilities and generalization offered by deep learning architectures have made them a popular choice within natural language text processing [143, 144]. The research presented in this chapter puts forward a sarcasm detection model for bilingual code-switch Hinglish tweets. The proposed model is a hybrid of Bi-LSTM with a *softmax* attention layer and CNN for real-time sarcasm detection. Feature engineering was done using pre-trained word embeddings for English, n-grams with TF-IDF encoding for Hindi hand-crafted features and frequency encoding for punctuations. Evaluation was carried out using different kinds of discrete baseline deep learning architectures for detecting sarcasm. The next section describes the related work in automatic sarcasm detection using deep learning architectures. The relevant studies focusing on code-mix and code-switch sarcasm detection are specifically included.

6.2 Related work

The mounting global interest of users in social media portals has reinforced analytics and sensing-based research areas to discover knowledge from the publicly available content. Tapping the opinion of users within this big pool of user-generated data has found many practical applications within the market and government intelligence domains. Twitter has been the top choice to mine content due to its global presence and accessibility. 'SA' [103] on all modalities (text, image, video, audio) of social data has been reported in literature. Primary studies with lexicon, ML and hybrid approaches are abundantly available. Literature is well-equipped with reviews and surveys on SA [5, 145, 137]. Pertinent studies specify contrast in opinion polarity to characterize sarcasm.

Deep learning models have been popularly used in NLP owing to their hierarchical learning and generalization capabilities [146]. Young et al. [123] discuss recent trends in deep learning based NLP. For detection of emotions, hybrid of attention based BiLSTM and CNN was explored by Felbo et al. [124]. Amir et al. [128] proposed a CNN trained using user and utterance based embeddings. Joshi et al. [147] used word embedding similarities as training features to detect sarcasm. Ghosh et al. [148] also worked on a hybrid of CNN, RNN and DNN. To explore syntactic and semantic information over tweets Zhang et al. [126] used Bi-LSTM in their work. Poria et al. explored different set of features viz. emotions; sentiment and personality based in their study [125]. Sarcastic sentiment detection has also been studied extensively. Mukherjee et al. [149] applied linguistic styles of authors for sarcasm detection on NB and fuzzy clustering algorithms. Suhaimin et al. [150] proposed a feature extraction process to detect sarcasm in bilingual texts. They used different categories of like feature lexical, pragmatic, prosodic and syntactic. The usage of context with neural network models for sarcasm detection was explored by Ren et al. [151]. They concluded that context-augmented neural model effectively decodes sarcastic clues from contextual information. Hazarika et al. [152] proposed a contextual sarcasm detector that adopts a hybrid approach of both content and context-driven modelling for sarcasm detection. They also explored usage of stylometric and personality features of the users in their work. Majumder et al. [153], presented a multitask learning-based framework using a deep neural network for sarcasm detection. Cai et al. [154], proposed a hierarchical fusion model to deal with the problem of multimodal sarcasm detection. They treated text features, image features and image attributes as three modalities to address this task.

Recently, SA in code-mixed languages has attracted much attention owing to the linguistic democratization and deep digital language divide. Sharma et al. [155] included abbreviations, word play, misspelled words and slang words and transliterated them to Romanized English words and analyzed the sentiment of the sentence using lexicon approach. The authors used text normalization before SA on FIRE 2013 and FIRE 2014 data which consists of English and Hindi languages to calculate final sentiment score by lexicon lookup in respective dictionaries. Joshi et al. [156] developed Hindi-SentiWordNet, a lexical resource to comprehend sentiments in Devanagari script. A fuzzy logic based approach for polarity detection in Hindi text was given by Rana [157]. Mittal et al. [158] also used Hindi SentiWordNet to classify sentiments using negation and discourse relation. Sharma et al. [159] presented various methods for SA of Hinglish language using the dataset of FIRE 2013 and FIRE 2014. Vyas et al. [160] explored POS tagging by using pure Hindi or English tagger and universal tagger in a code-mix script of Hindi and English in their research. Joshi et al. [161] introduced learning sub-word level representations in LSTM architecture for SA of Hindi-English code-mixed text. Bhargava et al. [162] proposed SA from code mixed sentences for English with combination of four different Indian languages. Vilares et al. [163] presented SA for English and Spanish in different environments. Malgaonkar et al. [164] worked on classifying the emotions into various categories like sad, sarcasm, happy, fear, bliss etc using sentences written in a combination of various languages. Konate et al. in [165] used six deep learning models for the SA task of code mixed Bambra-French language using the Facebook comments. Patra et al. [141] presented a paper for the task of sentiment identification from Hindi-English and Bengali-English code mixed datasets using the word and character level n- grams as features and SVM for the classification purpose. Abbasi et al. [166] also worked on a code mixed language of English and Arabic language and analyzed sentiment using entropy weighted genetic algorithms. Pakray et al.[167] addressed the difficulties in POS tagging of code-mixed data and introduced a Hidden Markov Model (HMM) based supervised algorithm as the solution of the same. Sane et al.[168] used deep learning techniques for humor detection in Hindi-English code-mixed tweets. Two most recent studies have focused on code-mix English-Hindi datasets for Sarcasm Detection [169], Irony detection [170] and Hate Speech Detection [171].

6.3 Data acquisition and preprocessing

6.3.1 Data acquisition

Real-time mash-up tweets were scrapped from Twitter using the Twitter Python API. We have mined the tweets using #sarcasm, keywords ‘sarcasm’ and ‘sarcastic’ and various the trending political (#government) and entertainment (#cricket, #bollywood) hashtags. Sarcasm Detector tool was used to build a randomly sampled dataset of 30000 tweets (12,000 sarcastic and 18000 non-sarcastic tweets). Out of these, the Hinglish mash-up tweets were extracted manually to create a balanced dataset of 3000 sarcastic and 3000 non-sarcastic tweets and considered as the dataset for this research.

6.3.2 Data pre-processing

Post data acquisition, data cleaning was done. The primary intent of pre-processing was to transform the data for extraction of features [172]. The process included:

- Removing numeric and empty texts, URLs, mentions, hashtags, stop-words and punctuations.
- Tokenization of tweets was done using the TreebankWordTokenizer of Python Natural Language Toolkit (NLTK) to filter the words, symbols and other elements called tokens [92]. The tokens were converted to lower case. For example, in a mash-up tweet “*Well done Aussies!! एक और हार LOL 👋👎*”, the tokens were generated as follows:

well	don	aussie	!	!	ए	औ	हार	lol	👋	👎

- The slangs and emojis were replaced by their descriptive text using the *SMS Dictionary* and *emojipedia* respectively.

well	don	aussie	!	!	ए	औ	हार	laugh	out	lou
clapping	hand	thumb	down							

- After tokenization, word-level language identification was performed using an online language identifier. This was primarily done to annotate each word with a source language tag <HINDI> or <ENGLISH> so that the words can be sent to the respective language processing module. Table 6.1 depicts this annotation for the example tweet.

Table 6.1. Word language annotation

Token word	**Language tag**
well	<ENGLISH>
done	<ENGLISH>
aussies	<ENGLISH>
एक	<HINDI>
और	<HINDI>
हार	<HINDI>
Laugh	<ENGLISH>
Out	<ENGLISH>
Loud	<ENGLISH>
Clapping	<ENGLISH>
Hands	<ENGLISH>
thumbs	<ENGLISH>
Down	<ENGLISH>

- Punctuations were discarded during data pre-processing phase but in casual or informal writing such as text message or online posts, these are used as a technique to add emphasis to written text. Therefore, the count of each punctuation mark (!, ?, ., capitalization, ‘x’, “x”) was extracted as a pragmatic feature set to train the model.

6.4 Feature extraction

The extraction of feature for training the model was done separately based on the source language as follows:

6.4.1 English language feature extraction

Word embedding is the collective name for a set of language modelling and feature learning techniques in NLP where words or phrases from the vocabulary are mapped to vectors of real numbers. Embeddings capture the representation of the word in higher dimensional plane. Through embeddings, we create a vector representation of the word which is learned by understanding the context of words. There are many techniques to create word embeddings for the deep neural models, such as, one-hot encoding; TF-IDF encoding, Word2Vec and GloVe. We used , GloVe, which is a count-based model to capture the semantic relatedness of English words, in our work. It learns by constructing a co-occurrence matrix (words X context) that basically count how frequently a word appears in a context.

6.4.2 Hindi language feature extraction

The extraction of feature using the Hindi language tokens is a multi-step procedure where the tokens are firstly tagged based on the part-of-speech and then mapped to the Hindi-SentiWordNet [156] to generate polarity and polarity scores and create a HindiSenti feature vector. The details of this feature extraction procedure are as follows:

- All Hindi words were passed to a Hindi POS tagger which assigns tags as adjective, noun, verb, adverb, interjection, common noun, question words etc. to each word. Thus, for the example given, the tags were generated as एक : Noun और: Adverb हार: Noun, Verb
- For each POS-tagged Hindi word, the polarity and polarity score was then determined using the Hindi-SentiWordNet (H-SWN). The H-SWN has various fields like (pos tag, Id, positive score, negative score, and synonyms). It is a subjective lexicon developed by IIT Bombay [156]. It contains words with part of speech and 3 scores positive, negative and objective. Sum of positive, negative and objective score sums to 1. The lexicon assigns single score to a word irrespective of the sense in which it is used. If the word is present in the file then we will compare the associated positive and negative scores to assign a value of +1 if the positive score is greater than the negative one or the value -1 assigned to that word representative of a negative sentiment(where negative score > positive score). If the word is not present in the file then the synonyms of the word are found and assigned the +1 or -1 value. Along with this an attribute that will store the polarity score ($PolScr$) of that word

is maintained. The polarity score within the range -1 to <0 shows negative sentiment and within the range >0 to 1 shows positive sentiment. A polarity score of 0 implies a neutral polarity. Thus, the polarity score ($PolScr$) is calculated as follows:

For a word with positive sentiment or (value=+1)
PolScr =+1 * positive score
whereas, for a word with negative sentiment or (value=-1)
PolScr = -1 * negative score

Following this approach the polarity score of every token of a tweet was calculated. A HindiSenti Feature Vector was then constructed by converting words into features using a language-independent n-grams technique with the *Tfidf* Vectorizer. The HindiSenti feature vector also kept the record of the corresponding polarity score for every feature.

6.4.3 Auxiliary pragmatic features

Whereas GloVe captures the conceptual relationship within words, pragmatic features necessarily portray the use of language. Punctuations such as exclamation mark, quotation marks, capatalization are used to add emphasis in written informal text and are significant signs which assist to comprehend the context inconsistency or intensity within the text as given below:

- Exclamation mark (!): Act as emotion intensifier without polarity shift. For example: "*She looks beautiful!!!!"* is more intense than "*She looks beautiful*."
- Wordplay: It involves word lengthening by repetition of alphabets. For example, the long sequence of word, "*sorryyyyy*" intensifies the expression.
- Uppercase word: Explicit use of uppercase alphabets intensifies the semantics. For example, "*I am so SORRY*!"is more persuasive than "*I am so sorry!"*

Moreover, the frequencies of occurrence of punctuations in a tweet strongly suggest sarcasm. Thus, frequency encoding was done for these pragmatic punctuation-based features to create a pragmatic feature vector p with six tuples, <R, E, I, P, U, Q>, where, R is the frequency of recurring alphabetic character, (that is, if recurrence > 2 set R =1, else 0) and E, I, P, U and Q defines the count of exclamation marks, question marks (interrogatory), periods, uppercase letters, single quotes('') or double quotes ("")

respectively. The conceptual flow of feature extraction in the proposed model is shown in fig. 6.1:

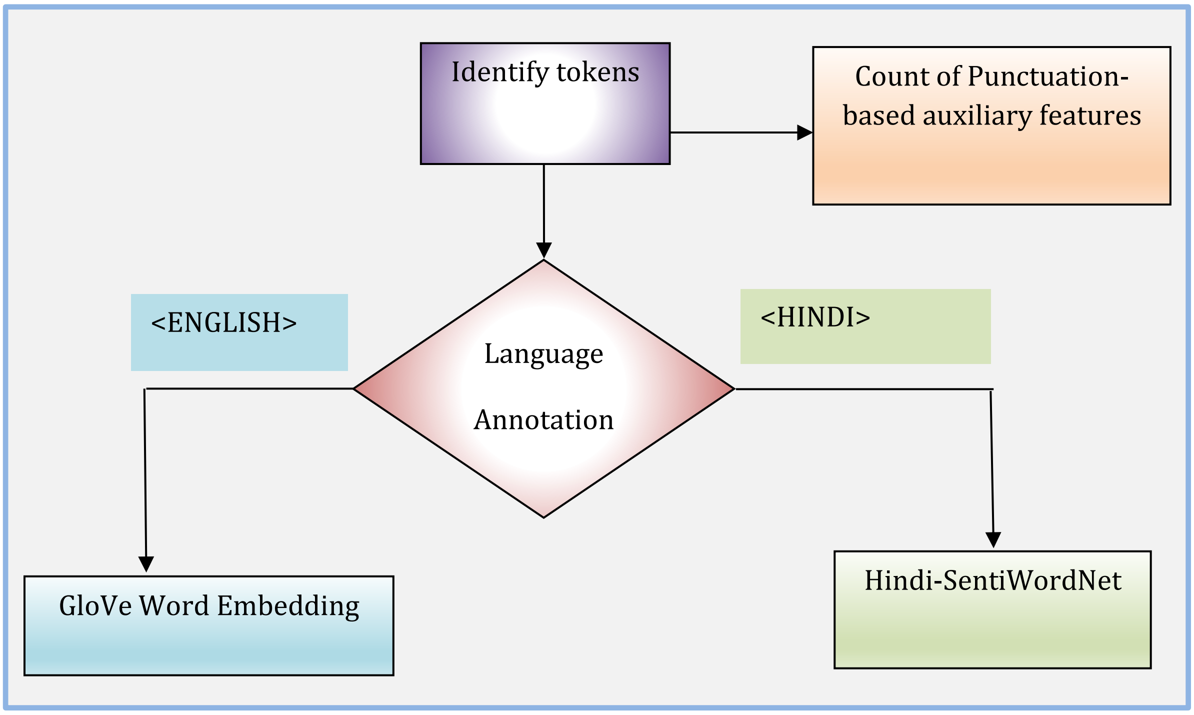


Fig. 6.1. Feature extraction in the proposed model

6.5 The proposed ^{softAtt}Bi-LSTM- $^{feature\text{-}rich}$CNN model

An automated model for detecting sarcasm in bi-lingual Hinglish mash-up tweets is a hybrid of Bi-LSTM with soft attention mechanism (^{softAtt}BiLSTM) and feature rich convolution neural network ($^{feature\text{-}rich}$CNN)trained using a combination of English, Hindi and auxiliary pragmatic feature vectors. Fig. 6.2 depicts the architecture of the proposed deep learning model.

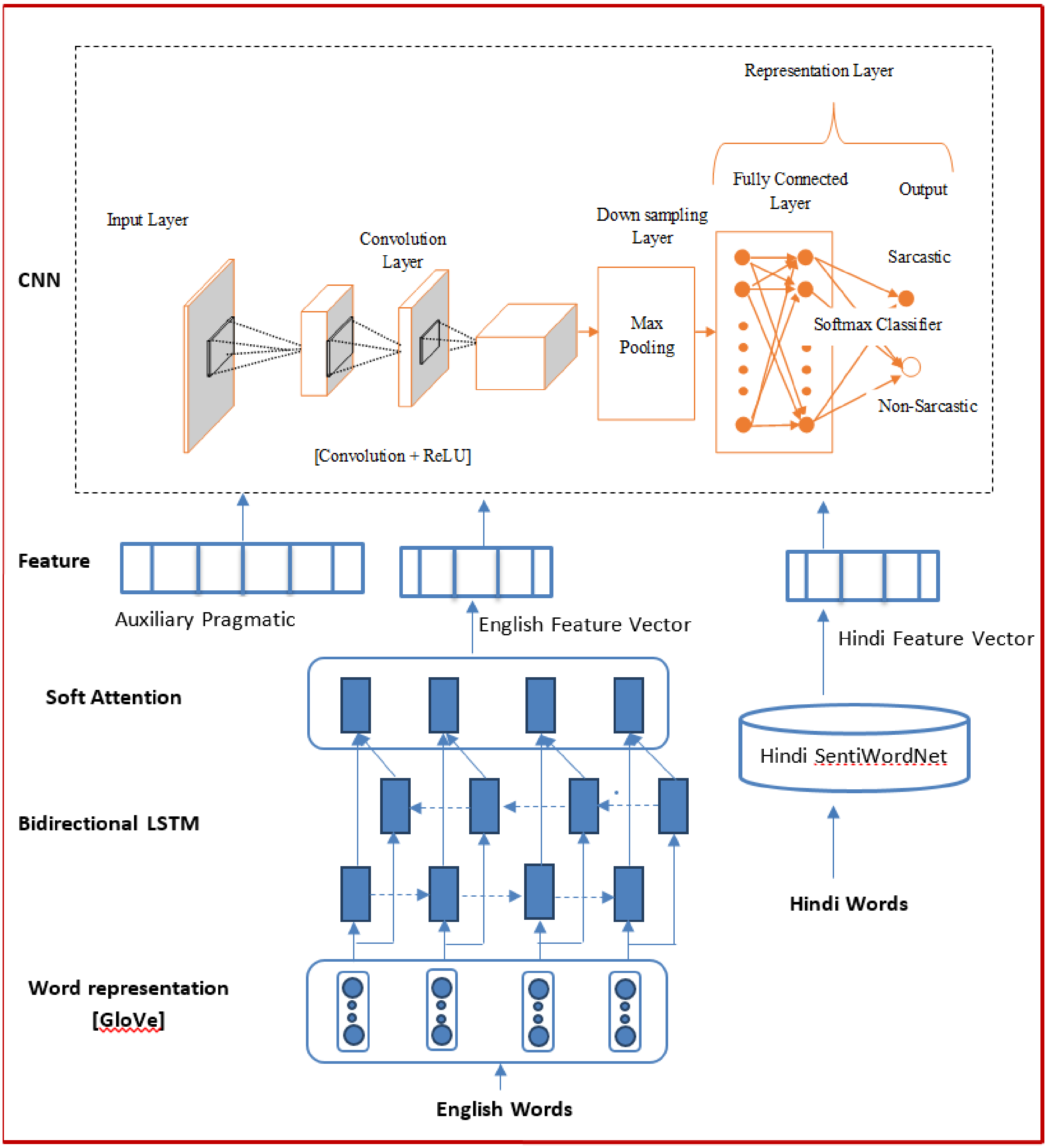


Fig.6.2. Architecture of the softAttBi-LSTM- feature-richCNN

The model consists of three modules, namely the English language processing module which generates context vectors using attention based Bi-LSTM; Hindi language processing module which generates TF-IDF encoded HindiSenti feature vector using Hindi-SentiWordNet (H-SWN); and the classifier module where CNN was trained using three feature vectors to generate the output predictions. The following sub-sections expound the details.

6.5.1 English context feature vector generation using Bi-LSTM with attention mechanism

This component consists of three layers of the deep learning architecture:

- **Embedding layer**: Embedding layer maps discrete words into lower dimensional vector space for computational efficiency. In this layer the English language input was mapped to real-valued vectors using pre-trained word embedding. The benefits of extracting features based on learned representations for sarcasm detection are available in literature [92].GloVe was used to generate 'word vector table' with an embedding dimension of 300 and a batch size of 50.
- **Bi-LSTM layer**: The LSTM in its core, preserves information from inputs that has already passed through it using the hidden state. Unidirectional LSTM only preserves information of the past because the only inputs it has seen are from the past. During training, errors back-propagate to adjust weights while maintaining the training properties. The calculation for the forget, the input and the output gate along with the input cell state was done using (1) to (6):

$$i_t = \sigma\,(W_i \cdot [h_{t-1}\,, x_t] + b_i) \qquad (1)$$

$$f_t = \sigma\,(W_f \cdot [h_{t-1}\,, x_t] + b_f) \qquad (2)$$

$$o_t = \sigma\,(W_o \cdot [h_{t-1}\,, x_t] + b_o) \qquad (3)$$

$$s_t = tanh\ (W_s \cdot [h_{t-1}\,, x_t] + b_s) \qquad (4)$$

$$c_t = f_t * c_{t-1} + i_t * s_t \qquad (5)$$

$$h_t = tanh\,(c_t) * \ o_t \qquad (6)$$

where,

- x_t is the t-th word vector that is it denotes the word representation of w_t
- $*$ represents the product (element wise)
- $W_i W_f W_o W_s$ are model parameters
- $b_i b_f b_o b_s$ represents the bias vectors

- The σ is the sigmoid function used as the gate activation function
- *tan h* is the hyperbolic tangent function

We have used Bi-LSTM to obtain word features H = (h_1, h_2, ..., h_n) concatenated from both directions. Bi-LSTM has two networks, one which reads information in forward direction and another which reads it in the reverse direction. This way the output is generated from both the past and future context. A forward LSTM processes the sentence (tweet) from x_1 to x_n, while a backward LSTM processes from x_n to x_1. For word x_t, a forward LSTM obtains a word feature as $\overrightarrow{h_t}$ and a backward LSTM obtains the feature as $\overleftarrow{h_t}$. Then, h is calculated using (7) [134]:

$$h_i = \left(\overrightarrow{h_i} \odot \overleftarrow{h_i}\right) \qquad (7)$$

Where, h_i is the output of the *i*-th word

$\odot$ function is a concatenation function used to combine the two outputs. Generally, different merge modes can be used to combine the outcomes of the Bi-LSTM layers. These are concatenation (default), multiplication, average, and sum.

$\vec{h}$, represents the output sequence of the forward layer which is calculated iteratively using inputs in a positive sequence from time t-n to time t-1,

$\overleftarrow{h}$ represents the output sequence of the backward layer which is calculated using the reversed inputs from time t-n to t-1.

- ***Attention layer***: The concept of attention is based on the hypothesis that words in a sentence (tweet/posts) have different importance quotient while defining the meaning in 'context'. In this work, the differentiable and deterministic *soft*-attention mechanism was used to generate the output as a weighted combination of all the input states rather than just last state (Fig.6.3).

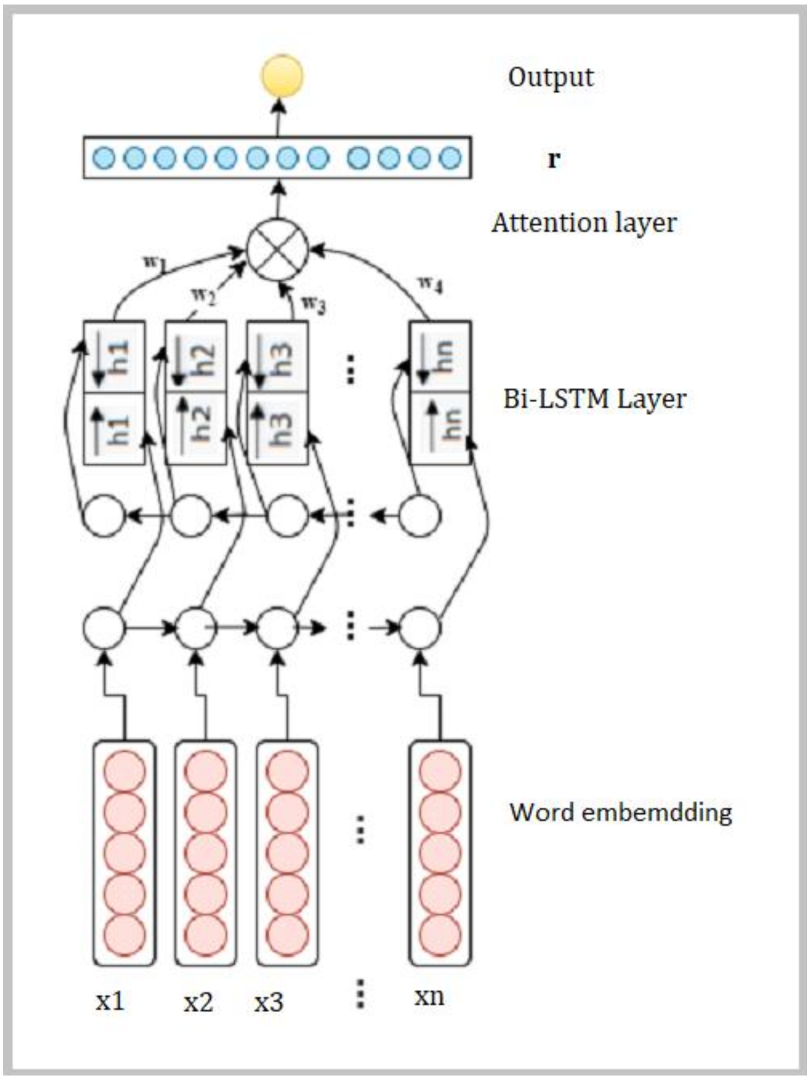


Fig.6.3. Bi-LSTM with attention

Using this word-level attention mechanism [173], the attention score e_i was given to each word i in the sentence t as given by "(8)".

$$e_i = tanh(W_h h_i + b_h) \qquad (8)$$

Where W_h and b_h are the weight and bias from the attention layer.

And then using attention mechanism, the weight w_i was assigned to each word feature h_i using "(9)".

$$w_i = \frac{\exp(e_i)}{\sum_{t=1}^{N} \exp(e_t)} \quad , \quad \sum_{i=1}^{N} w_i = 1 \qquad (9)$$

The hidden states were finally calculated to produce a hidden sentence feature vector r by a weighted sum function using “(10)”.

$$r = \sum_{i=1}^{N} w_i \, h_i \qquad (10)$$

6.5.2 Feature-rich CNN

The CNN was enriched with features from both languages which define the mash-up along with pragmatic marker features. That is, the input to the CNN is a combination of the output features of the $^{\text{sofAtt}}$Bi-LSTM, auxiliary punctuation-based features and the HindiSenti feature vector. It comprises the next three layers of the deep learning architecture, that is, the convolution layer, the down-sampling layer and the representation layer (Fig. 6.4):

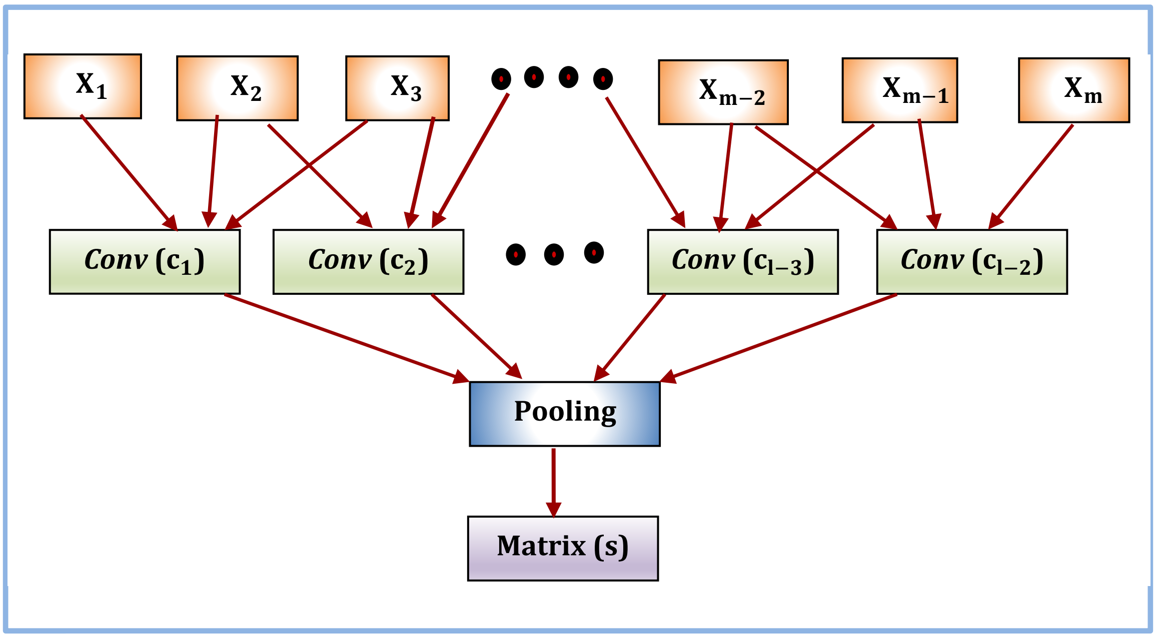


Fig.6.4. Convolution and pooling

- **Convolution layer:** Filters perform convolutions and generate (variable-length) feature maps. The model used three layer convolution architecture with a total of 100 convolution filters each for window size (3, 3). The dropout regularization was set to 0.5 to ensure that the model does not overfit. ReLU activation function was applied to introduce nonlinearities into the model and generate a rectified feature map.
- **Down-sampling layer:** Also known as pooling layer, this layer down samples the feature map and retains the non-trivial features. In this study, max pooling was used, in order capture the maximum value within the rectified feature map.

- **Representation layer:** This is the output layer which consists of a dense layer where every node in the layer is connected to every node in the preceding layer thus forming a fully-connected layer. This layer classifies the output using a *softmax* function to generate a value between 0 (non-sarcastic) and 1 (sarcastic).

6.6 Results

The discussion of results is done in two parts: (i) performance of the proposed deep learning model (ii) performance g.

6.6.1 Performance results

The predictive modelling using 6000 random Hinglish tweets was done. The performance was evaluated using accuracy, recall, precision, and F-measure. The following Table 6.2 depicts the results of the proposed ^{softAtt}Bi-LSTM- $^{feature\text{-}rich}$CNN deep learning model.

Table 6.2. Performance of the ^{softAtt}Bi-LSTM- $^{feature\text{-}rich}$CNN model

Performance Quantifier	Value (%)
Accuracy	92.71
Recall	90.67
Precision	89.49
F-measure	89.05

6.6.2 Comparison with baselines

The model was evaluated individually for different language modules and hand-crafted features. The results of using CNN for Hindi words and CNN for Hindi words with punctuations are shown in tables 6.3 and 6.4 respectively.

Table 6.3. Performance of CNN-Hindi language

Performance Quantifier	Value (%)
Accuracy	43.44
Recall	42.78
Precision	42.89
F-measure	45.63

Table 6.4. Performance of CNN-Hindi language with punctuations

Performance Quantifier	Value (%)
Accuracy	54.94
Recall	50.22
Precision	53.9
F-measure	54.64

The performance for English language with and without punctuations was evaluated for both CNN and Bi-LSTM models individually. The following tables 6.5, 6.6, 6.7 and 6.8 depict the results.

Table 6.5. Performance of CNN-English language

Performance Quantifier	Value (%)
Accuracy	70.35
Recall	69.92
Precision	70.59
F-measure	71.12

Table 6.6. Performance of CNN-English language with punctuations

Performance Quantifier	Value (%)
Accuracy	81.28
Recall	80.61
Precision	83.01
F-measure	82.72

Table 6.7. Performance of Bi-LSTM-English language

Performance Quantifier	Value (%)
Accuracy	69.75
Recall	68.28
Precision	64.91
F-measure	66.22

Table 6.8. Performance of Bi-LSTM-English language with punctuations

Performance Quantifier	Value (%)
Accuracy	71.03
Recall	70.27
Precision	68.77
F-measure	69.45

The graph in fig. 6.5 depicts the comparison of accuracies of deep models for individual languages with and without punctuation

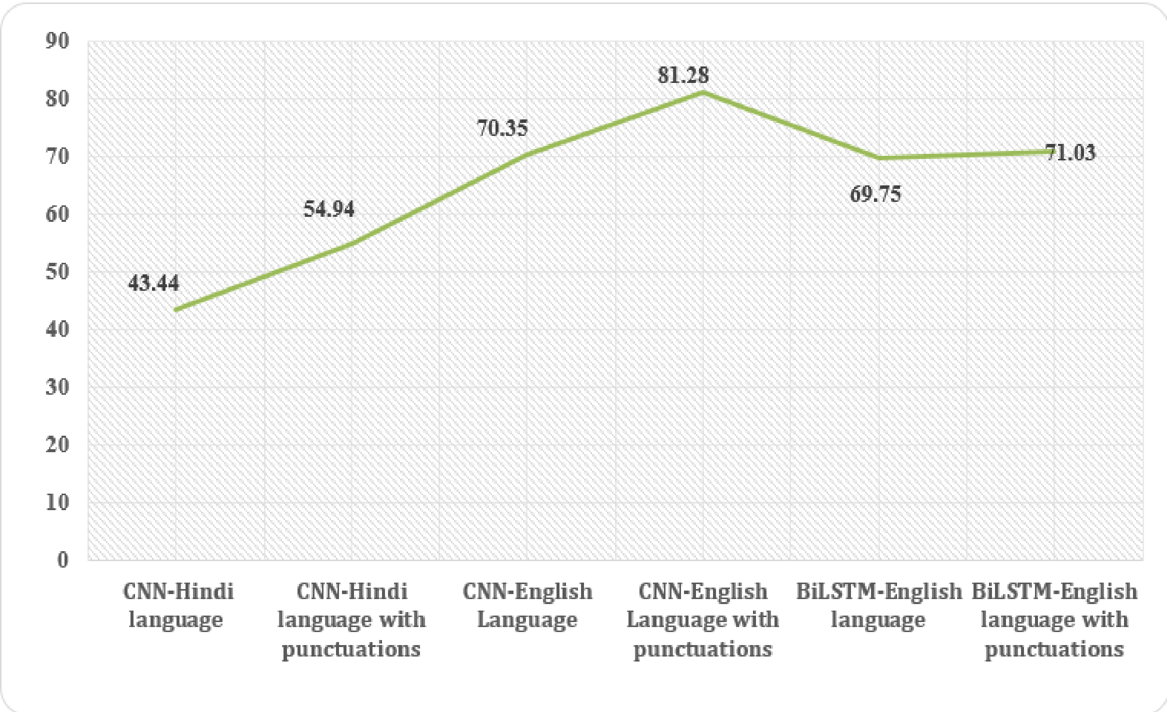


Fig.6.5. Comparison of individual deep learning model on language modules

The proposed model was also compared with two deep learning hybrid architectures, namely, LSTM with $^{\text{feature-rich}}$CNN and Bi-LSTM without attention hybrid with $^{\text{feature-rich}}$CNN. Table 6.9 and 6.10 presents the results. The graph in fig.6.6 presents the comparison of these two hybrid deep learning models to the proposed model.

- LSTM-feature-richCNN

Table 6.9. Performance of the LSTM-feature-richCNN model

Performance quantifier	Value (%)
Accuracy	81.75
Recall	78.28
Precision	74.91
F-measure	76.45

- Bi-LSTM-feature-richCNN (Bi-LSTM without attention)

Table 6.10. Performance of the Bi-LSTM-feature-richCNN model

Performance quantifier	Value (%)
Accuracy	85.03
Recall	81.27
Precision	78.77
F-measure	79.51

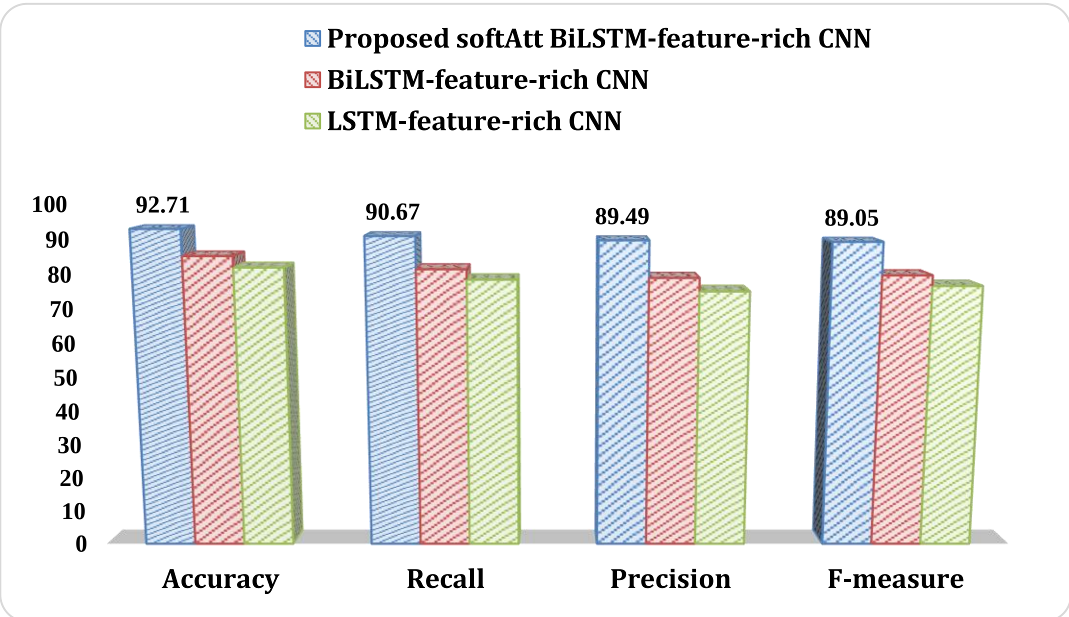


Fig.6.6. Comparison of hybrid deep learning model performance

The results demonstrated that the proposed softAttBi-LSTM- feature-richCNNmodel achieved the highest accuracy of 92.71% as compared to the other variants. The order of models from highest to lowest accuracy is, softAttBi-LSTM- feature-richCNN>Bi-LSTM- feature-richCNN>LSTM- feature-richCNN. Fig.6.7 which depicts the comparison of F-measure obtained for each model. The proposed softAttBi-LSTM- feature-richCNN also achieved the highest recall, precision and F-measure.

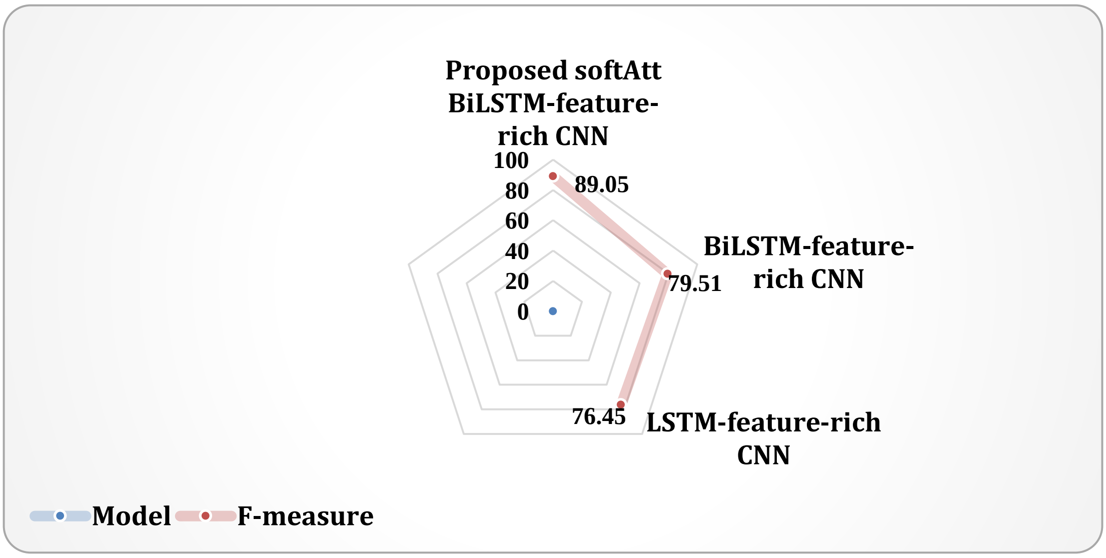


Fig.6.7. F-measure of the proposed model with baselines

6.7 Chapter summary

Detecting sarcasm, irony and emotions in real-time, code-switched user-generated text is imperative for fine-grained SA. This chapter presented a foremost learning model for real-time sarcasm detection in Hinglish code-switch dataset. The *softmax* attention based Bi-LSTM learned semantic context vector for English features from the GloVe word embeddings and forwarded it to the CNN. The HindiSenti feature vector and auxiliary punctuation-based features were also combined with the English feature vector to train the CNN. Performance benchmarking was done to evaluate the effectiveness of the proposed model and high classification accuracy and F-measure were observed. The model outperformed the baseline deep learning models with a superior classification accuracy of 92.71% and F-measure of 89.05%. Mining sentiments in Hindi is a challenge as it is a morphologically rich and a free order language as compared to English. This amplifies the intricacies of managing user-generated content. The sarcastic tone classification was thus limited by the capabilities of Hindi-SentiWordNet for extraction of hand-crafted Hindi language features.

CHAPTER 7

Application of context for sarcasm detection in multimodal data

The work presented in this chapter attempts to detect sarcasm in multimodal data. A model as a sarcastic meme predictor for analyzing sarcasm from visual language of Instagram memes is presented . Section 7.1 discusses the introduction to sarcasm and memes. Section 7.2 presents the related work carried on sarcasm detection. Section 7.3 presents the model developed for detecting sarcasm in multimodal data. Section 7.4 discusses the results and chapter winds up with summary (section 7.5).

7.1 Introduction

Recently, as memes and GIFs dominate the social feeds; typo-graphic visual content has become a considerable element of social media. 'Meme' is a viral image or video often altered by internet users for humorous effect. These convey human expressions but with a wide range of emotions, and often require context to fully understand humor and sarcasm. "Sarcasm is a type of sentiment where people express their negative feelings using positive or intensified positive words in the text" [102]. It is an expression representative of conflict between the apparent and the applied. Memes are topic-dependent and highly contextual, therefore, polarity shift and other contextual clues can help detect sarcasm from text and improve the generic sentiment classification of typo-graphic social data. For example, as in the post shown in Fig. 7.1, "Being nice" demonstrates a conflict between the obvious state of "hardest part of the job". This inconsistency, contrast and shifts within the polarities of sentiments validate sarcasm as a distinctive case of SA.

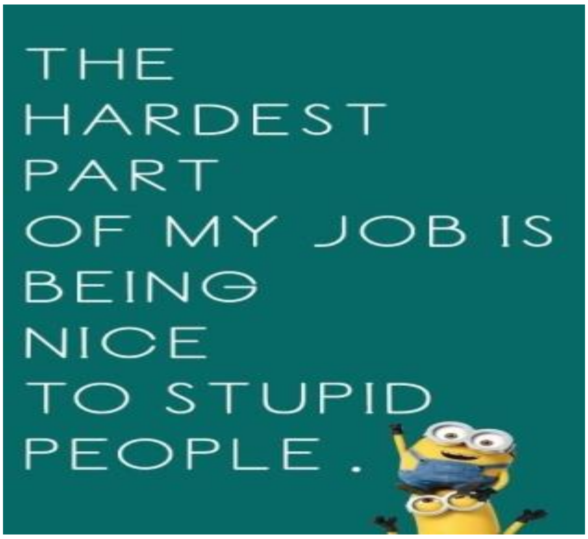


Fig.7.1. Example of sarcastic meme

Detecting sarcastic tone is very difficult to accomplish without having a sufficient knowledge of the 'context' of the situation, the specific topic, and the environment[174]. Textual SA has been widely studied [5,19,175,176]; few related studies have also reported visual analysis of images to predict sentiment [89,177,178] but the domain of visual text which combines both text and image has been least explored in literature. This combination can be observed in two ways as follows:

- ***Typo-graphic:*** Artistic way of text representation (Fig. 7.2 & 7.3). It is primarily textual content but in an image form using variety of fonts and styling options to change the apperance of text.

Fig.7.2. Example of sarcastic typo-graphic meme

- ***Info-graphic**:* Text embedded along with an image (Fig.7.3). The image may either intensify the textual expresssion or contradict the written content.

Fig.7.3. Example of sarcastic info-graphic meme

We developed a model to deal with sarcasm detection in typographic memes. Next section presents the literature work carried on sarcasm detection in memes.

7.2 Related work

Textual sarcasm detection has been studied by researchers as a specialized case of sentiment classification. Both ML [100,101, 104,105] and deep learning [179,180] models have been explored to predict sarcastic tone in online user-generated content, especially on twitter. The growth in use of images to express opinion online makes text-based SA and subsequently sarcasm detection restricted in terms of capturing the sentiment associated. Image-based SA has emerged as a significant research domain with the research in this field falling under three areas which are: aesthetics [181-183], emotion detection [85-87, 184-190] and sentiment ontology [84]. To the best of our knowledge no work on visual language based typo-graphic images has been done. This work is an attempt to capture the sarcastic tone in typo-graphic memes.

7.3 The proposed sarcastic meme predictor model

Automatic sarcasm detection as a typical classification task has the primary goal to distinguish between sarcastic or non-sarcastic content. This research probed the problem of detecting sarcasm in memes from the computational linguistic perspective by relying on lexical, pragmatic and semantic cues. Context was defined using various features such as n-grams with TF-IDF encoding, frequency encoding for punctuations, sentiment based hand-crafted features. These linguistic context markers were used to train and test classifiers which predict sarcasm effectively and efficiently in the MemeBank dataset built. The following figure 7.4 depicts the architecture of the proposed Sarc-M model for saracsm detection in tyopographic memes.

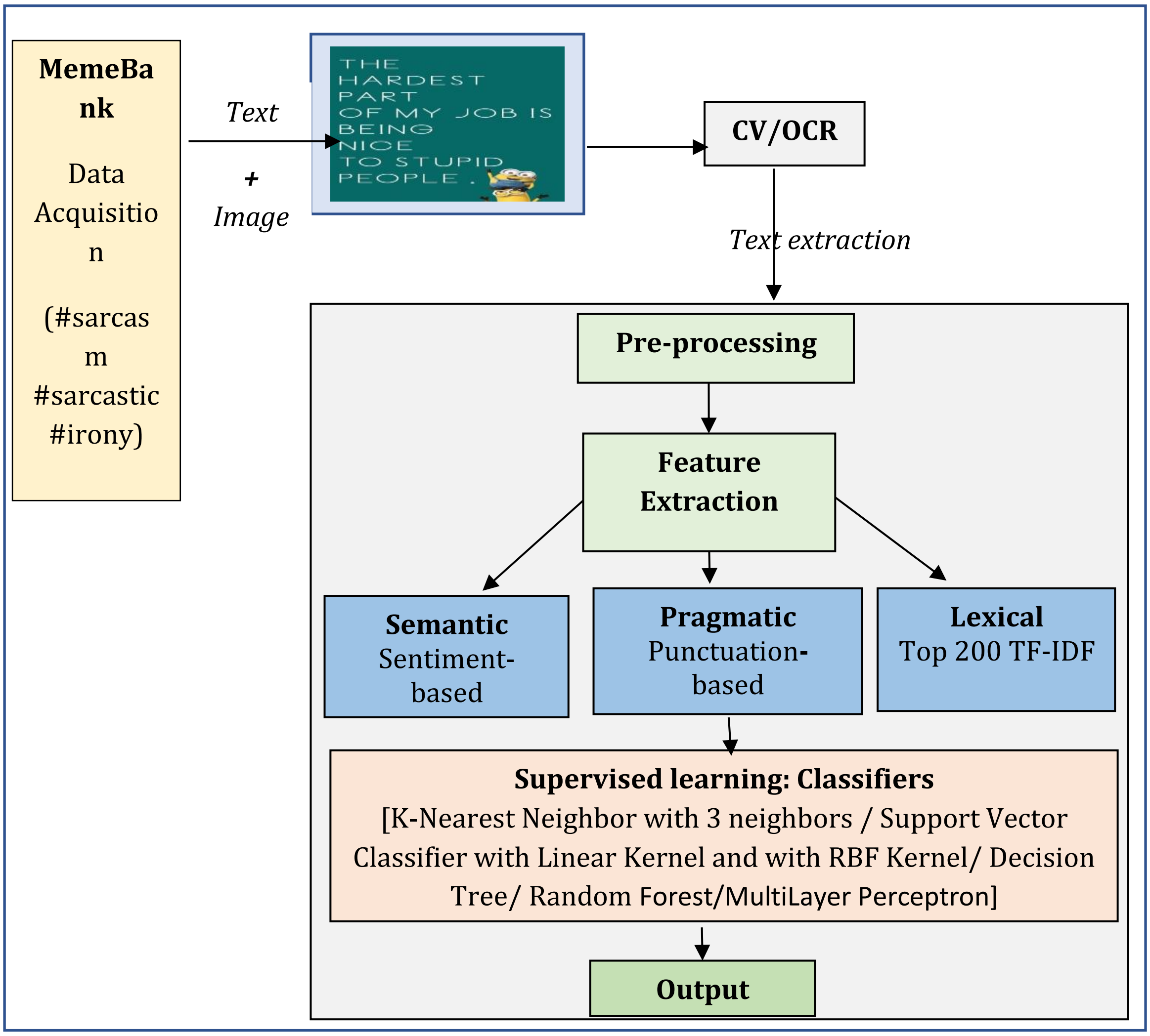


Fig.7.4. Architecture of the proposed Sarc-M model

The following sub-sections present the details:

7.3.1 Data acquisition

The visual language corpus, MemeBank, was constructed using images from Instagram. This MemeBank is a balanced dataset consisting of 1200 images, which includes600 typo-graphic images collectedwith hashtags #sarcastic, #sarcasm and #irony and negative class data of 600 images collected using hashtags #motivational and #inspirational.

7.3.2 Text extraction

For typo-graphic text, we used the CV API to extract text using OCR from the image. We used Open CV version 3.4.2, EAST text detector proposed by Zhou et.al, 2017 [94], which is a deep learning model, based on a novel architecture and training pattern. It is capable of running at near real-time at 13 FPS on 720p images and obtains state-of-the-art text detection accuracy. The extracted text was passed through an OCR for recognition. The following figure 7.5 shows the sample text extraction using the API.

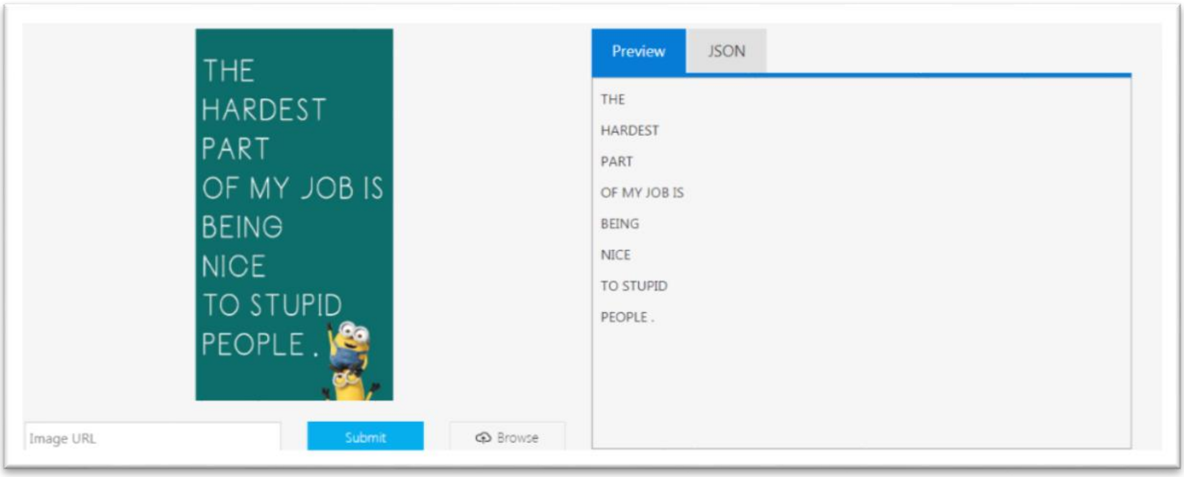


Fig.7.5. Sample text extraction using the API

7.3.3 Pre-processing

Post text extraction, pre-processing was done to clean and transform the text for extraction of features. This pre-processing included:

- Removal of stop words, placeholders, mentions etc.
- Replacement of URLs, special characters such as @, #.
- Use of Natural Language tool-Kit (NLTK) for tokenization.
- Use of Porter's stemmer for stemming to the root word.

- Removal of non-ASCII English character.
- Part-of- Speech tagging was also done to extract common structural patterns such as verb, adverb, adjective and noun.
- Although all the punctuation marks in a tweet were removed as the part of the cleaning process; however, the count of each punctuation mark was kept as we use them as pragmatic features to train the model. In this work, five punctuation-based features that represent figurative text and provide symbolic clues within the tweet were used. These include exclamation marks (!), question marks (?), periods (.), capital letters and use of "or".

7.3.4 Feature extraction

Three categories of features, namely, lexical, pragmatic and semantic features were used in our work.

- Firstly, the conventional lexicon based statistical weighing term frequency-inverse document frequency (TF-IDF) [8] measure was used and the top 200 entries were filtered.
- Secondly, frequency of punctuation-based pragmatic features was used. Punctuations, wordplay or uppercase alphabets characterize symbolic clues which help comprehend the context within the text. Frequency encoding was done for these pragmatic features as shown in the following table 7.1 [64]

Table 7.1. Pragmatic features

Feature	Decription
rep	frequency of repetitive alphabets, that is, if alphabet repetition > 2, then set feature to true else false
excl	frequency of exclamation marks
ques	frequency of question marks
dots	frequency of dots
caps	frequency of capital letters
quotes	frequency of '' or ""

- Thirdly, while pragmatic features are concerned with the language use, sentiment-based semantic features typically convey the conceptual relationship between words. The sarcastic sentences are skewed in the sense that the sentiment polarity of the literal meaning may differ radically from the intended figurative meaning. The shift in sentiment (positive to negative or negative to positive) within these sentences is a strong indicator of sarcasm. For example, in the sentence "I love everybody. Some I love to be around, some I love to avoid and others I would love to punch on their face" clearly expresses sentiment polarity shifts. Therefore, any type of inconsistency or contradiction between sentiments within the text can be considered hinting sarcasm. The sentiment-based features defined in a previous reported work [64] were used in this research. The authors characterize two lists of words classified as "positive words" and "negative words" which contain words with the positive emotional content i.e. emotional positive terms (e.g., "love", "enjoy", "happy", etc.) and that with negative emotional content i.e. emotional negative terms (e.g., "hate", "worry", "sad", etc.). These lists of words were created using the pos_tag library under NLTK [92]. Next, using these lists, for each textual content the number of positive words (pw) and negative words (nw) were counted. Also, the adjectives, adverbs and verbs have higher emotional content as compared to nouns, therefore, all the words, either positive or negative that have the associated POS tag, were counted another time to create two additional features that represent the number of highly emotional positive terms (PW) and highly emotional negative terms(NW) respectively. Sometimes, emotional content is conveyed through Hashtags too. For example, in the textual post, "Thanks a lot for always helping me #ihateyou", the hashtag"#ihateyou" tells that the user is not actually thanking the addressed user, but was rather extremely disliking him for not helping him. In addition to the already mentioned features, some features that are related to the contrast between the sentimental components were also extracted. Contrast means the co-existence of both a negative and a positive component in the same tweet. The ratio of emotional words was calculated denoted as $\rho(t)$ expressed in the following equation 1:

$$\rho(t) = ((\delta \cdot PW + pw) - (\delta \cdot NW - nw)) / ((\delta \cdot PW + pw) + (\delta \cdot NW - nw)) \quad (1)$$

Where,

pw is the count of words with positive sentiment

nw is the count of words with negative sentiment

PW is the count of words with highly positive emotional content

NW is the count of words with highly negative emotional content

ρ is the score determined to find contrast between the above sentimental components

t is the tweet

δ is a weight bigger than 1 given to the highly emotional words and is set to 3.

The sentiment related features considered for this study are summarized in Table 7.2.

Table 7.2. Sentiment-based features

Feature	Description
Pw	Count of words with positive sentiment
Nw	Count of words with negative sentiment
PW	Count of words with highly emotional positive content
NW	Count of words with highly emotional negative content
ρ score	Finds contrast between the above sentimental components, here delta is equal to 3.

7.3.5 Supervised learning

Classifiers refer to the type of task-specific algorithms where the feature extraction is problem specific and the model is trained using the handcrafted features. The model extracts 11 features (sentiment +punctuation) which were concatenated with the top TF-IDF features generated. This feature matrix was used to train the five baseline classifier namely, K-Nearest Neighbor (KNN), SVM, RF, DT and MLP. We considered KNN with 3 neighbors and the support vector classifier with both Linear and RBF Kernel. The description of these techniques is available in relevant literature studies[13,19,71,191].

7.4 Results

This section illustrates the performance achieved by the sarcasm classifier based on accuracy and F1 score. The following table 7.3 illustrates the performance results achieved. It was observed from the above results that the MLP gives the highest accuracy for the dataset, that is, 87.95 % followed by the RF classifier. The least accuracy of 73.25% was observed using SVC with RBF kernel.

Table 7.3. Performance Results of the proposed model

Performance measure	KNN with 3 neighbors	SVC with linear kernel	SVC with RBF kernel	Decision tree	Random forest	Multi layer perceptron
Accuracy	76.12	73.25	78.34	79.37	80.35	87.95
F1 score	71.58	80.19	72.67	70.70	80.56	80.09

The figures 7.6 and 7.7 depict the results graphically.

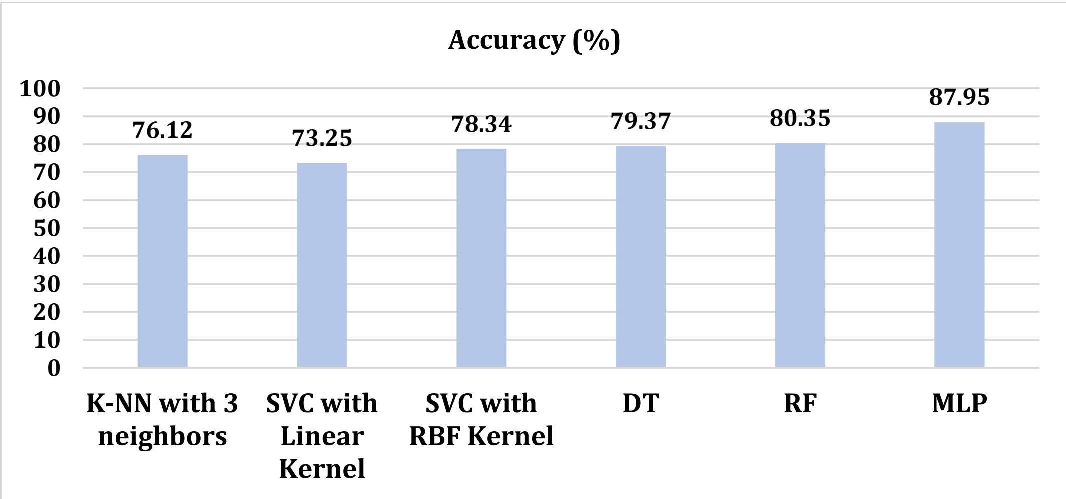


Fig.7.6. Accuracy achieved by various classifiers

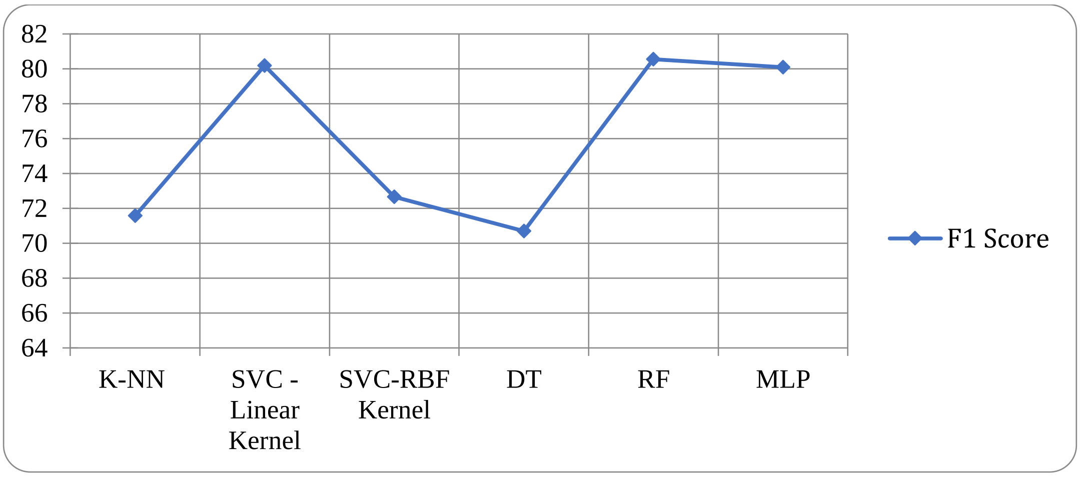


Fig.7.7. F1 Score of various classifiers

As observed from fig.7.7, the F1 score of 80.56 has been reported for RF classifier.

7.5 Chapter summary

Visual language defines the current trend of expression on social platforms and memes further underlines this growing internet culture. Automatic saracsm detection in text has been a dynamic area of research. This chapter presented a model to detect sarcasm in typo-graphic memes from the visual language of Instagram. A MemeBank containing 1200 typo-graphic images was built from Instagram. CV API was used to extract text for analytics from the typo-graphic memes. A combination of hand-crafted features was used to train and test five different classifiers namely, KNN, SVM, RF, DT and MLP. It was observed that the best sarcasm classification was achieved by MLP whereas the least was observed for SVC with linear kernel.

CHAPTER 8

Conclusion & future work

This chapter presents the summary of work carried out in this research in the area of CBSA. Section 8.1 presents the research summary followed by listing of limitations of the work in section 8.2. A thorough discussion of future scope and open areas of the research have been discussed in section 8.3. And finally, conclusion of the thesis is presented in section 8.4.

8.1 Research summary

The research presented in this thesis started with the aim to find an answer for capturing and using contextual information to improve upon the task of SA which may assist fine-grain SA, emotion analysis, sarcasm detection, irony detection, humor detection, among others. Based on the literature review carried out in this domain, it was found that traditional SA shows its limiting performance accuracy when exposed to non-standard and ambiguous language usage. Mining the massive, high-dimensional, noisy, incomplete social data presents new challenges for SA. The knowledge of appropriate context of the content under review can help in resolving these issues and improve accuracy of prediction. Context is a set of facts or circumstances that surround a situation or event. Understanding context is one of the most difficult aspects of content moderation.

Studies reveal that context is a multifaceted concept with no standard categorizations. Our work intends to formalize the concept of context in SA by defining types of contextual cues which may assist fine-grain SA, emotion analysis, sarcasm detection, irony detection, humor detection, among others. A multifaceted concept of context has been described in detail. Based upon the knowledge of context, a model incorporating contextual information for carrying SA of textual data was proposed in this research. Currently there exists two main types of approaches for sentiment detection namely, ML and lexicon-based. The proposed model was a hybrid of the two approaches and works at two levels. At first, a ML based method was used to classify the tweet into one of the three positive, negative or neutral category. At the second level, a lexicon based approach, which captures context related information, was used for finding sentiment polarity and strength of the tweet. Then polarity and strength of the two levels were combined to assign one of the seven classes to the given tweet ranging from high positive, moderate positive, low positive to neutral to low negative, moderate negative and high negative. In addition to this work, the contextual framework work dealing with single modality (textual data) was extended to deal with multiple modality data. Images are more expressive than text and at the same time text embedded or represented as an image further defines this power of expressiveness. This research proposed a model for SA to capture this expressiveness for text in an image, both typographic or infographic, as sentiment polarity and strength. The model developed for multimodal SA model of Twitter can serve as a visual listening tool for enhanced social media monitoring and analytics. The performance results were motivating and improved the generic SA task.

The conceptualization of 'context' was further applied for detection of specialized sentiment like sarcasm with improved accuracy. This research probed the problem of detecting sarcasm in social media from the computational linguistic perspective by relying on lexical, pragmatic and semantic cues. Context was defined using various features such as n-grams with TF-IDF encoding, frequency encoding for punctuations, hand-crafted features and pre-trained word embedding. Three models were proposed using these linguistic context markers to train and test shallow as well as deep classifiers which predict sarcasm effectively and efficiently in two experimental datasets (SemEval 2015 Task 11 and 20k posts from Reddit). As a step further, the work presented in this research, focused on sarcasm detection in the mash-up language (literal code-switch) tweets. Multilinguality is a well-recognized challenge of NLP. The cultural diversities, country-specific trending topics and hash-tags on social media and easy availability of native language keyboards for social media applications add to the variety and volume of user-generated content in diverse languages and dialects. The proposed model is a hybrid of Bi-LSTM with a softmax attention layer and CNN for real-time sarcasm detection in Hinglish, which refers to the juxtaposition of words from Hindi and English language. Lastly, a model Sarc-M, for sarcasm detection in typo-graphic memes (multimodal data) using supervised learning based on lexical, pragmatic and semantic features was developed. The learning model was evaluated using five different classifiers and the results were evaluated using a balanced dataset of typo-graphic images, called MemeBank, scrapped from Instagram.

8.2 Limitation of study

Following are some limitations of our work which restricts improvement in contextual models:

- Benchmark datasets for multimodal and multilingual data analysis are not available.
- Our work includes extraction of text from image, which is limited by extraction capability of Optical character reader. The OCR has a limited capability for handwritten text recognition and suffers from reduced accuracy with 'lack of contrast' images where the text color and the background color are almost similar.
- The sarcastic tone classification in Hinglish data is limited by the capabilities of Hindi-SentiWordNet for extraction of hand-crafted Hindi language features. Mining sentiments in Hindi is a challenge as it is a morphologically rich and a free order language as compared to English. This amplifies the intricacies of managing user-generated content.
- The model for mash-up language relies on the efficiency of online language identifier and part-of-speech tagger which have not yet reached the state-of-the-art accuracy.

8.3 Future directions

- Benchmark datasets for multimodal and multilingual data are required to be developed for research.
- There is a need for developing lexicons for Hindi language.
- In this research, only the text and image modality type were considered whereas other modalities such as animated GIFs and memes define an open problem within the research domain.
- The use of word embeddings other than GloVe like FastTex, Bert etc. and deep learning for context-aware SA of text can be explored.
- Identifying context, performing word sense disambiguation and wordplay are open research problems associated with both English and Hindi languages.

- The use of code-mix and code-switch languages, intentional ambiguity and novelty in vocabulary, throw new challenges in sarcastic tone detection in short text, making it a dynamic area of research.
- The current work gives promising results on sarcasm detection in short text as contained in the Tweets, there occurs a lot more scope with respect to sarcasm detection for longer texts.
- This work can be extended by including more contextual information related to the history of users' tweets. Including details about a user's social network may also help in detecting sarcasm as a person is more likely to use sarcasm in friendly circles than in formal settings.
- A further enhancement in the accuracy can be tested by applying other computing techniques like fuzzy logic, swarm intelligence algorithms and deep learning approaches for detection of sarcasm.
- Apart from Twitter , other social media like Facebook , Instagram etc. can also be explored more for analyzing sarcasm from the texts, images, videos etc.

8.4 Conclusion

Automated text analytics has many promising practical applications and improvement in the same facilitate good decision making. Social media serves as a communication channel and a social listening, awareness, activism & feedback tool for stakeholder engagement and cooperation. Keeping tabs on social media sentiment, that is, *'the online mood'* is a key part of social media listening. Determining the literal exactitude and opinion polarity in text is imperative for real-time SA. SA task attains fitting results with literal language where articulation is same as anticipated interpretation. The non-standard vocabulary with informal textual content and noise currently define the content strategy of social media. Based upon the literature review conducted, it was revealed that the knowledge of appropriate context of the content under review can be used for resolving these issues. The developed contextual model for carrying SA of textual data shows best performance with accuracy of 87.37% with topical data(#Demonetization). This result demonstrated the role of context in improving the accuracy of classification. The contextual framework dealing with single modality (textual data) when extended to deal with multiple modality data(both typographic as well infographic) shows 91.32 %

performance accuracy. This result validated that how context improves the generic SA task while simultaneously handling multimodal data issue. The excellent performance results makes this model, a visual listening tool for enhanced social media monitoring and analytics. As an application of contextual SA, this research probed the problem of detecting implicit sentiments like sarcasm in social media. The empirical analysis was carried out using lexical, pragmatic and semantic cues in varied combinations, two datasets on three different models which uses classifiers ranging from shallow to deeper ones. It was observed that the Bi-LSTM using pre-trained vectors from the GloVe word embedding is the best sarcasm prediction model with an average accuracy of approximately 85% (~84.6). The result of this analysis validated that application of context helps in improving the accuracy of sarcasm detection task. While dealing with the task of multilingual sarcasm detection , the hybrid model of Bi-LSTM with a softmax attention layer and CNN achieves the highest accuracy of 92.71% as compared to state of art hybrid models i.e. LSTM-$^{\text{feature-rich}}$CNN model and Bi-LSTM-$^{\text{feature-rich}}$CNN model. The results of Sarc-M model for typo-graphic memes (multimodal data) using supervised learning based on lexical, pragmatic and semantic features demonstrated the importance of context for sarcasm detection.

The aim of this research i.e. to find an answer for capturing and using contextual information to improve upon the task of SA which may assist fine-grain SA, emotion analysis, sarcasm detection, irony detection, humor detection, among others has been fulfilled successfully. The models proposed in this research proved to perform better in comparison to various state-of-art techniques. Overall, the results obtained using deep context analysis were motivating and successfully tackled the NLP-based issues in conventional SA. Based upon observations and findings, we conclude that context plays an important part in social media text mining. CBSA helps to improve upon the limitation of traditional SA techniques by considering contextual peculiarities and sarcasm contamination. It gives an additional dimension to the SA task and paves the way for a proper understanding of the tone and sense of the post.

References

[1] Cambria, Erik, et al., eds. "A practical guide to sentiment analysis." Cham, Switzerland: Springer International Publishing, (2017).
[2] Eke, Christopher Ifeanyi, et al. "Sarcasm identification in textual data: systematic review, research challenges and open directions." Artificial Intelligence Review 53.6 (2020): 4215-4258.
[3] Barman U, Das A., Wagner J., Foster J. "Code Mixing: A Challenge for Language Identification in the Language of Social Media", EMNLP (2014), pp. 13.
[4] Nilep C. "Code switching in sociocultural linguistics", Colorado Research in Linguistics, vol. 19, no. 1(2006), pp. 1-22.
[5] Pang B., Lee L. "Opinion mining and sentiment analysis, Foundations and trends in information retrieval", vol. 2, no. 1–2(2008), 1-135.
[6] Deng, Li, and Yang Liu, eds. "Deep learning in natural language processing". Springer, (2018).
[7] Bhatia, M. P. S., and Akshi Kumar. "Paradigm shifts: from pre-web information systems to recent Web-based contextual information retrieval." Webology 7, no. 1 (2010).
[8] Bhatia, MPS. & Kumar, A. (2008). Information retrieval and machine learning: supporting technologies for Web mining research and practice. Webology, 5(2), article55.
[9] Liu Bing. "Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing", Second edition, 2010
[10] Wiebe, Janyce, Wilson, Theresa, Bruce, Rebecca, Bell, Matthew, and Martin, Melanie "Learning subjective language". Computational Linguistics, 2004, 30(3):277–308
[11] Pang, Bo and Lee Lillian. "Opinion mining and sentiment analysis". Foundations and Trends in Information Retrieval 2008,(1-2),1–135
[12] Kumar, Akshi, and Mary Sebastian Teeja. "Sentiment analysis: A perspective on its past, present and future." International Journal of Intelligent Systems and Applications 4, no. 10 (2012): 1.
[13] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal 5, no. 4 (2014): 1093-1113.
[14] Kumar, Akshi, and Teeja Mary Sebastian. "Sentiment analysis on twitter." IJCSI International Journal of Computer Science Issues 9, no. 4 (2012): 372.
[15] Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." In Proceedings of the conference on human language technology and empirical methods in natural language processing, pp. 347-354. Association for Computational Linguistics, 2005.
[16] Das, Sanjiv, and Mike Chen. "Yahoo! for Amazon: Extracting market sentiment from stock message boards." In Proceedings of the Asia Pacific finance association annual conference (APFA), vol. 35, p. 43. 2001.
[17] Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." In Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417-424. Association for Computational Linguistics, 2002.
[18] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pp. 79-86. Association for Computational Linguistics, 2002.
[19] Kumar, Akshi, and Arunima Jaiswal. "Empirical Study of Twitter and Tumblr for Sentiment Analysis using Soft Computing Techniques." In Proceedings of the World

Congress on Engineering and Computer Science, vol. 1. 2017.
[20] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, “Systematic Literature Reviews in Software Engineering- A systematic literature review”, Information and Software Technology, Elsevier, Vol.51(1):7-15, 10.1016/j.infsof.2008.09.009.
[21] Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. by Jo Akers. Paperback, 292 Pages, Published 2009. *ISBN*-10: 1-900640-47-3 / 1900640473.
[22] Han, H., Bai, X. & Li, P. Neural Comput & Applic (2018). https://doi.org/10.1007/s00521-018-3698-4
[23] Majumder, N., Hazarika, D., Gelbukh, A., Cambria, E., & Poria, S. (2018). Multimodal sentiment analysis using hierarchical fusion with context modeling. Knowledge-Based Systems, 161, 124-133.
[24] Sheik, Reshma, Sebin S. Philip, Arjun Sajeev, Sinith Sreenivasan, and Georgin Jose. "Entity Level Contextual Sentiment Detection of Topic Sensitive Influential Twitterers Using SentiCircles." In Data Engineering and Intelligent Computing, pp. 207-216. Springer, Singapore, 2018.
[25] Feng, S., Wang, Y., Liu, L. et al. World Wide Web (2018). https://doi.org/10.1007/s11280-018-0529-6
[26] Vechtomova, Olga. "Disambiguating context-dependent polarity of words: An information retrieval approach." Information Processing & Management 53, no. 5 (2017): 1062-1079.
[27] Hung, Chihli. "Word of mouth quality classification based on contextual sentiment lexicons." Information Processing & Management 53, no. 4 (2017): 751-763.
[28] Deng, Shuyuan, Atish P. Sinha, and Huimin Zhao. "Resolving Ambiguity in Sentiment Classification: The Role of Dependency Features." ACM Transactions on Management Information Systems (TMIS) 8, no. 2-3 (2017): 4.
[29] Tao, Wan, and Tao Liu. "Building ontology for different emotional contexts and multilingual environment in opinion mining." Intelligent Automation & Soft Computing (2017): 1-7.
[30] Saif, Hassan, Miriam Fernandez, Leon Kastler, and Harith Alani. "Sentiment lexicon adaptation with context and semantics for the social Web." Semantic Web 8, no. 5 (2017): 643-665.
[31] Zhou, Yunxiao, Man Lan, and Yuanbin Wu. "ECNU at SemEval-2017 Task 4: Evaluating Effective Features on Machine Learning Methods for Twitter Message Polarity Classification." In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pp. 812-816. 2017.
[32] Jiménez-Zafra, Salud María, Arturo Montejo-Ráez, Maite Martin, and L. Alfonso Urena Lopez. "SINAI at SemEval-2017 Task 4: User based classification." In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pp. 634-639. 2017.
[33] Fersini, Elisabetta, Federico Alberto Pozzi, and Enza Messina. "Approval network: a novel approach for sentiment analysis in social networks." World Wide Web 20, no. 4 (2017): 831-854.
[34] Muhammad, Aminu, Nirmalie Wiratunga, and Robert Lothian. "Contextual sentiment analysis for social media genres." Knowledge-based systems 108 (2016): 92-101.
[35] Saif, Hassan, Yulan He, Miriam Fernandez, and Harith Alani. "Contextual semantics for sentiment analysis of Twitter." Information Processing & Management 52, no. 1 (2016): 5-19.
[36] Meire, Matthijs, Michel Ballings, and Dirk Van den Poel. "The added value of auxiliary data in sentiment analysis of Facebook posts." Decision Support Systems 89 (2016): 98-112.
[37] Nakov, Preslav, Sara Rosenthal, Svetlana Kiritchenko, Saif M. Mohammad, Zornitsa Kozareva, Alan Ritter, Veselin Stoyanov, and Xiaodan Zhu. "Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts." Language Resources and Evaluation 50, no. 1 (2016): 35-65.

[38] Frankenstein W., Joseph K., Carley K.M. (2016) Contextual Sentiment Analysis. In: Xu K., Reitter D., Lee D., Osgood N. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2016. Lecture Notes in Computer Science, vol 9708. Springer, Cham
[39] Wu, Fangzhao, Yongfeng Huang, and Yangqiu Song. "Structured microblog sentiment classification via social context regularization." Neurocomputing 175 (2016): 599-609.
[40] Gaspar, Rui, Cláudia Pedro, Panos Panagiotopoulos, and Beate Seibt. "Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events." Computers in Human Behavior 56 (2016): 179-191.
[41] Gelli, Francesco, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. "Image popularity prediction in social media using sentiment and context features." In Proceedings of the 23rd ACM international conference on Multimedia, pp. 907-910. ACM, 2015.
[42] Jurek, Anna, Maurice D. Mulvenna, and Yaxin Bi. "Improved lexicon-based sentiment analysis for social media analytics." Security Informatics 4, no. 1 (2015): 9.
[43] Hridoy, Syed Akib Anwar, M.Tahmid Ekram, Mohammad Samiul Islam, Faysal Ahmed, and Rashedur M. Rahman. "Localized twitter opinion mining using sentiment analysis." Decision Analytics 2, no. 1 (2015): 8.
[44] Dragoni, Mauro, Andrea GB Tettamanzi, and Célia da Costa Pereira. "Propagating and aggregating fuzzy polarities for concept-level sentiment analysis." Cognitive Computation 7, no. 2 (2015): 186-197.
[45] Recupero, Diego Reforgiato, Valentina Presutti, Sergio Consoli, Aldo Gangemi, and Andrea Giovanni Nuzzolese. "Sentilo: frame-based sentiment analysis." Cognitive Computation 7, no. 2 (2015): 211-225.
[46] Poria, Soujanya, Erik Cambria, Gregoire Winterstein, and Guang-Bin Huang. "Sentic patterns: Dependency-based rules for concept-level sentiment analysis." Knowledge-Based Systems 69 (2014): 45-63.
[47] Li, Yung-Ming, Lienfa Lin, and Shih-Wen Chiu. "Enhancing targeted advertising with social context endorsement." International Journal of Electronic Commerce 19, no. 1 (2014): 99-128.
[48] Lau, Raymond YK, Chunping Li, and Stephen SY Liao. "Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis." Decision Support Systems 65 (2014): 80-94.
[49] Yang, Bishan, and Claire Cardie. "Context-aware Learning for Sentence-level Sentiment Analysis with Posterior Regularization." In ACL (1), pp. 325-335. 2014
[50] Anjaria, Malhar, and Ram Mohana Reddy Guddeti. "A novel sentiment analysis of social networks using supervised learning." Social Network Analysis and Mining 4, no. 1 (2014): 181.
[51] Andrea Vanzo, Danilo Croce, Roberto Basili (2014): A context based model for Sentiment Analysis in Twitter. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics, pp. 2345–2354, Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 2014.
[52] Korenek, Peter, and Marián Šimko. "Sentiment analysis on microblog utilizing appraisal theory." World Wide Web 17, no. 4 (2014): 847-867.
[53] Ren, Fuji, and Ye Wu. "Predicting user-topic opinions in twitter with social and topical context." IEEE Transactions on Affective Computing 4, no. 4 (2013): 412-424.
[54] Bosco, Cristina, Viviana Patti, and Andrea Bolioli. "Developing corpora for sentiment analysis: The case of irony and senti-tut." IEEE Intelligent Systems 28, no. 2 (2013): 55-63.
[55] Weichselbraun, Albert, Stefan Gindl, and Arno Scharl. "Extracting and grounding contextualized sentiment lexicons." IEEE Intelligent Systems 28, no. 2 (2013): 39-46.
[56] Liu, Yang, Xiaohui Yu, Aijun An, and Xiangji Huang. "Riding the tide of sentiment change: sentiment analysis with evolving online reviews." World Wide Web 16, no. 4 (2013): 477-

496.
[57] Aisopos, Fotis, George Papadakis, Konstantinos Tserpes, and Theodora Varvarigou. "Content vs. context for sentiment analysis: a comparative analysis over microblogs." In Proceedings of the 23rd ACM conference on Hypertext and social media, pp. 187-196. ACM, 2012.
[58] Dey, Lipika, and SK Mirajul Haque. "Opinion mining from noisy text data." International Journal on Document Analysis and Recognition (IJDAR) 12, no. 3 (2009): 205-226.
[59] Sánchez-Rada, J. Fernando, and Carlos A. Iglesias. "Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison." Information Fusion 52 (2019): 344-356.
[60] Zeng, Jiangfeng, Xiao Ma, and Ke Zhou. "Enhancing attention-based LSTM with position context for aspect-level sentiment classification." IEEE Access 7 (2019): 20462-20471.
[61] Feng, Shi, et al. "Attention based hierarchical LSTM network for context-aware microblog sentiment classification." World Wide Web 22.1 (2019): 59-81.
[62] Zuo, Enguang, et al. "Context-Specific Heterogeneous Graph Convolutional Network for Implicit Sentiment Analysis." IEEE Access 8 (2020): 37967-37975.
[63] Phan, Huyen Trang, et al. "Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model." IEEE Access 8 (2020): 14630-14641.
[64] Bouazizi, M., & Ohtsuki, T. O.: A pattern-based approach for sarcasm detection on Twitter. IEEE Access, 4, 5477-5488 (2016).
[65] Bamman, D., & Smith, N. A.: Contextualized sarcasm detection on twitter. In Ninth International AAAI Conference on Web and Social Media, (2015).
[66] Wang, Z., Wu, Z., Wang, R., & Ren, Y.: Twitter sarcasm detection exploiting a context-based model. In International Conference on Web Information Systems Engineering (pp. 77-91). Springer, Cham (2015).
[67] Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H. and Demirbas, M., 2010, July. Short text classification in twitter to improve information filtering. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp.841-842).ACM.
[68] Kumar, A., Sebastian, T.M.,2012. Machine learning assisted sentiment analysis. In Proceedings of International Conference on Computer Science & Engineering, ICCSE (pp. 123-130).
[69] Joachims, T., 1998, April. Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning (pp. 137-142).Springer, Berlin, Heidelberg.
[70] Rokach, L. and Maimon, O., 2005. Top-down induction of decision trees classifiers-a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 35(4), pp.476-487.
[71] Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.
[72] Dietterich, T.G., 2000, June. Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15). Springer, Berlin, Heidelberg.
[73] Galar, M., Fernández, A., Barrenechea, E., Bustince, H. and Herrera, F., 2011. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. Pattern Recognition, 44(8), pp.1761-1776.
[74] Geurts, P., Ernst, D. and Wehenkel, L., 2006. Extremely randomized trees. Machine learning, 63(1), pp.3-42.
[75] Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan,R.P. and Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. Journal of chemical information and computer sciences, 43(6), pp.1947-1958.
[76] Freund, Y. and Schapire, R.E., 1997. A decision- theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), pp.119-

139.
[77] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, pp.1189-1232.
[78] Saif, H., Fernandez, M., He, Y. and Alani, H., 2014,May. Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In European Semantic Web Conference (pp. 83-98). Springer, Cham.
[79] Saif, H., He, Y., Fernandez, M. and Alani, H., 2014,October. Semantic patterns for sentiment analysis of Twitter. In International Semantic Web Conference (pp.324-340). Springer, Cham.
[80] Da Silva, N.F., Hruschka, E.R. and Hruschka Jr, E.R., 2014.Tweet sentiment analysis with classifier ensembles.Decision Support Systems, 66, pp.170-179.
[81] Esuli, A. and Sebastiani, F., 2007. SentiWordNet: a high-coverage lexical resource for opinion mining. Evaluation,17(1), p.26.
[82] Scott A. Golder and Michael W. Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and day length across diverse cultures. Science 333, 6051 (2011), 1878–1881.
[83] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. 2011. Twitter mood predicts the stock market. Journal of Computational Science 2 (2011), 1–8.
[84] Borth, D., Chen, T., Ji, R. and Chang, S.F., 2013, October. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In Proceedings of the 21st ACM international conference on Multimedia (pp. 459-460). ACM.
[85] Zhao S, Yao H, Wang F, Jiang X, Zhang W. Emotion based image musicalization. In Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on 2014 Jul 14 (pp. 1-6). IEEE.
[86] Zhao S, Yao H, Yang Y, Zhang Y. Affective image retrieval via multi-graph learning. In Proceedings of the 22nd ACM international conference on Multimedia 2014 Nov 3 (pp. 1025-1028). ACM.
[87] Zhao S, Gao Y, Jiang X, Yao H, Chua TS, Sun X. Exploring principles-of-art features for image emotion recognition. In Proceedings of the 22nd ACM international conference on Multimedia 2014 Nov 3 (pp. 47-56). ACM.
[88] Wang Y, Wang S, Tang J, Liu H, Li B. Unsupervised Sentiment Analysis for Social Media Images. In IJCAI 2015 Jul 25 (pp. 2378-2379).
[89] Katsurai M, Satoh SI. Image sentiment analysis using latent correlations among visual, textual, and sentiment views. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on 2016 Mar 20 (pp. 2837-2841). IEEE.
[90] Chen T, Salah Eldeen HM, He X, Kan MY, Lu D. VELDA: Relating an Image Tweet's Text and Images. In AAAI 2015 Jan 25 (pp. 30-36).
[91] Mandhyani J, Khatri L, Ludhrani V, Nagdev R, Sahu S. Image Sentiment Analysis. International Journal of Engineering Science. 2017 Feb;4566.
[92] Bird S, Loper E. NLTK: the natural language toolkit. In Proceedings of the ACL 2004 on Interactive poster and demonstration sessions 2004 Jul 21 (p. 31). Association for Computational Linguistics.
[93] Porter MF. An algorithm for suffix stripping. Program. 1980 Mar 1;14(3):130-7.
[94] Zhou X, Yao C, Wen H, Wang Y, Zhou S, He W, Liang J. EAST: an efficient and accurate scene text detector. In Proc. CVPR 2017 Jul 21 (pp. 2642-2651).
[95] Pak A, Paroubek P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Int. Lang. Resour. Eval. 2010; 1320–1326.
[96] Liu B. (2015). Sentiment Analysis Mining Opinions, Sentiments, and Emotions. Cambridge University Press. 2015.
[97] Chen J, Yan S, Wong K C. (2018). Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis.Neural Computing and Applications.1-10.
[98] Zubiaga A., Aker A., Bontcheva K., Liakata M., Procter, R.(2017).Detection and resolution

of rumours in social media: A survey. arXiv preprint arXiv:1704.00656.
[99] Satapathy R., Guerreiro, C., Chaturvedi, I., Cambria, E. (2017). Phonetic-Based Microtext Normalization for Twitter Sentiment Analysis, in: ICDM, 407– 413.
[100] Joshi A, Bhattacharyya P, Carman M. (2017). Automatic Sarcasm Detection: A Survey. ACM Computing Survey. Article No. 73, Volume 50 Issue 5.
[101] Hercig T, Lenc L. (2017). The Impact of Figurative Language on Sentiment Analysis. Proceedings of the International Conference Recent Advances in Natural Language Processing. 301–308.
[102] Bharti S.K., Vachha B., Pradhan R K., Babu K S, Jena S K. (2016). "Sarcastic sentiment detection in tweets streamed in real time: a big data approach", Digital Communications and Networks. vol. 2, no. 3, pp. 108-21.
[103] Dave K, Lawrence S, Pennock DM. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web. ACM. 519-528.
[104] Camp E. (2012). Sarcasm, pretense, and the semantics/pragmatics distinction. Noûs. 46(4):587-634.
[105] Riloff E, Qadir A, Surve P, Silva LD, Gilbert N, Huang R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. EMNLP. 13:704–714.
[106] Davidov D, Tsur O, Rappoport A. (2010).Semi-supervised recognition of sarcastic sentences in twitter and Amazon. Proceedings of the fourteenth conference on computational natural language learning, Association for Computational Linguistics. 107-116.
[107] González-Ibánez R, Muresan S, Wacholder N. (2011).Identifying sarcasm in Twitter: a closer look. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers Association for Computational Linguistics. 2: 581-586.
[108] Liebrecht CC, Kunneman F A, van Den Bosch APJ.(2013). The perfect solution for detecting sarcasm in tweets# not.
[109] Liu P, Chen W, Ou G, Wang T, Yang D, Lei K. (2014). Sarcasm detection in social media based on imbalanced classification. International Conference on Web-Age Information Management. SpringerCham 2014. 459-471.
[110] Barbieri F, Saggion H, Ronzano F. (2014). Modelling sarcasm in twitter, a novel approach. Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 50-58.
[111] Ptáček T, Habernal I, Hong J. (2014). Sarcasm detection on Czech and English twitter. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 213-223.
[112] Bharti SK, Babu KS, Jena SK. (2015). Parsing-based sarcasm sentiment recognition in twitter data. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM .1373-1380.
[113] Rajadesingan A, Zafarani R, Liu H. (2015).Sarcasm detection on twitter: A behavioral modeling approach. Proceedings of the Eighth ACM International Conference on Web Search and Data, ACM. 97-106.
[114] Sulis E, Farías DI, Rosso P, Patti V, Ruffo G. (2016).Figurative messages and affect in twitter: Differences between# irony,# sarcasm and# not. Knowledge-Based Systems. 15;108:132-43.
[115] Karoui J, Benamara F, Moriceau V, Aussenac-Gilles N, HadrichBelguith L. (2015).Towards a contextual pragmatic model to detect irony in tweets. Association for Computational Linguistics (ACL).
[116] Wang Shui-Hua, Muhammad, Khan & Hong, Jin& Kumar, Arun & Zhang, Yu-Dong. Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization, Neural Computing and Applications, 2019, doi:

10.1007/s00521-018-3924-0
[117] Wang,Shui-Hua Junding Sun, Preetha Phillips, Guihu Zhao, Yudong Zhang: Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units, Journal of Real-Time Image Processing, 2018, 15(3): 631-642
[118] Tang Chaosheng, Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling, Frontiers in Neuroscience, 2018, 12, Article ID: 818
[119] Zhang, Y.D, Chichun Pan, Junding Sun, Chaosheng Tang. (2018). Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU, Journal of Computational Science, 2018, 28: 1-10
[120] Zhang, Y.D, Yin Zhang, Xiao-Xia Hou, Hong Chen, Shuihua Wang. (2018). Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed. Multimedia Tools and Applications. 2018, 77(9): 10521-10538
[121] Zhang, Y.D, Chichun Pan, Xianqing Chen, Fubin Wang. (2018). Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling, Journal of Computational Science, 2018, 27: 57-68
[122] Zhang, Y.D, Khan Muhammad, Chaosheng Tang (2018) Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform, Multimedia Tools and Applications, 2018, 77(17): 22821-22839
[123] Young T, Hazarika D, Poria S, Cambria E. (2018). "Recent trends in deep learning based natural language processing". IEEE Computational Intelligence Magazine 13 (3), 55-75.
[124] Felbo B., Mislove A., Søgaard A., Rahwan I., Lehmann S. (2017). "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm", arXiv preprint arXiv:1708.00524
[125] Poria S, Cambria E, Hazarika D, Vij P. (2016). A deeper look into sarcastic tweets using deep convolutional neural networks. Proc. Int. Conf. Computational Linguistics, pp. 1601-1612
[126] Zhang M, Zhang Y, Fu G. (2016). Tweet sarcasm detection using deep neural network. Proceedings of COLING 2016, The 26th International Conference on Computational Linguistics: Technical Papers. 2449-2460.
[127] Joshi A, Tripathi V, Patel K, Bhattacharyya P, Carman M. (2016). Are Word Embedding-based Features Useful for Sarcasm Detection? arXiv preprint arXiv:1610.00883.
[128] Amir S, Wallace BC, Lyu H, Silva PC. (2016). Modelling context with user embeddings for sarcasm detection in social media. arXiv preprint arXiv:1607.00976.
[129] Ghosh A, Li G, Veale T, Rosso P, Shutova E, Barnden J, Reyes A. (2015). Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). 470-478.
[130] Pennington J, Socher R, Manning C. (2014).Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 1532-1543.
[131] Hltcoe J. (2013). Semeval-2013 task 2: Sentiment analysis in Twitter. Atlanta, Georgia, USA, 312.2013.
[132] Khodak M, Saunshi N, Vodrahalli K. (2017). A large self-annotated corpus for sarcasm. arXivpreprint arXiv:1704.05579.
[133] Salehinejad H, BaarbeJ, Sankar S, Barfett J, Colak E, Valaee S. (2017). Recent Advances in Recurrent Neural Networks. arXiv preprint arXiv:1801.01078.
[134] Graves A, Jaitly N, Mohamed AR. (2013). Hybrid speech recognition with deep bidirectional LSTM. Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop2013. 273-278.
[135] Powers, D.M. (2011). "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", Journal of Machine Learning Technologies.

[136] Roy, A., & Todorovic, S. (2018, August). Learning to Learn Second-Order Back-Propagation for CNNs Using LSTMs. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 97-102). IEEE.
[137] Matras Y., Bakker P. (2003). "The mixed language debate: Theoretical and empirical advances", Walter de Gruyter, vol. 145.
[138] P. Arora, (2013). "Sentiment Amalysis for Hindi Language" Master thesis, IIT, Hyderabad.
[139] Zheng W., Qiang Y E. (2009). "Sentiment Classification of Chinese Traveler Reviews by Support Vector Machine Algorithm", IEEE Third International Symposium on Intelligent Information Technology Application.
[140] Mandal, S., Das, S D., Das D. (2018). "Language identification of Bengali-English code-mixed data using character & phonetic based lstm models". arXiv preprint arXiv:1803.03859.
[141] Patra, B G., Das, D., Das A. (2018). "Sentiment analysis of code-mixed Indian languages: An overview of sail code-mixed shared task@ icon-2017." arXiv preprint arXiv:1803.06745.
[142] Parshad R D., Bhowmick, S., Chand, V. Kumari, N., Sinha, N. (2016). "What is India speaking? Exploring the "Hinglish" invasion." Physica A: Statistical Mechanics and its Applications 449 (2016): 375-389.
[143] Sivanandam, S N., Deepa S N. (2007). "Principles of Soft Computing (With CD)", John Wiley & Sons.
[144] Lei, P., & Todorovic, S. (2016, October). Recurrent temporal deep field for semantic video labeling. In European Conference on Computer Vision (pp. 302-317). Springer, Cham.
[145] Wang, K., Bansal, M., & Frahm, J. M. (2018, March). Retweet wars: Tweet popularity prediction via dynamic multimodal regression. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1842-1851). IEEE.
[146] Wang, R., Frahm, J. M., & Pizer, S. M. (2018). Recurrent neural network for learning dense depth and ego-motion from video. arXiv preprint arXiv:1805.06558.
[147] Joshi A., Bhattacharyya P., Carman M., Saraswati J., Shukla R. (2016). "How do cultural differences impact the quality of sarcasm annotation?: A case study of indian annotators and american text.", In Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities pp. 95-99.
[148] Ghosh A., Veale T. (2016). "Fracking sarcasm using neural network", In Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp. 161-169, 2016.
[149] Mukherjee, S., & Bala, P. K. (2017). Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering. Technology in Society, 48, 19-27.
[150] Suhaimin, M. S. M., Hijazi, M. H. A., Alfred, R., & Coenen, F. (2017, May). Natural language processing based features for sarcasm detection: An investigation using bilingual social media texts. In 2017 8th International Conference on Information Technology (ICIT) (pp. 703-709). IEEE.
[151] Ren, Y., Ji, D., & Ren, H. (2018). Context-augmented convolutional neural networks for twitter sarcasm detection. Neurocomputing, 308, 1-7.
[152] Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., & Mihalcea, R. (2018). Cascade: Contextual sarcasm detection in online discussion forums. arXiv preprint arXiv:1805.06413.
[153] Majumder, N., Poria, S., Peng, H., Chhaya, N., Cambria, E., & Gelbukh, A. (2019). Sentiment and sarcasm classification with multitask learning. IEEE Intelligent Systems, 34(3), 38-43.
[154] Cai, Y., Cai, H., & Wan, X. (2019, July). Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 2506-2515).

[155] Sharma S, Srinivas PY, Balabantaray RC. (2015).”Text normalization of code mix and sentiment analysis”. In Advances in Computing, Communications and Informatics (ICACCI), 2015 pp. 1468-1473, IEEE.
[156] Das, A., & Bandyopadhyay, S. (2010, August). SentiWordNet for Indian languages. In Proceedings of the Eighth Workshop on Asian Language Resources, pp. 56-63.
[157] Rana S. (2014). “Sentiment Analysis for Hindi Text using Fuzzy Logic”. Indian Journal of Applied Research, Vol.4, Issue.8.
[158] Mittal, N., Agarwal B., Chouhan G., Bania N., Pareek P. (2013). “Sentiment Analysis of Hindi Review based on Negation and Discourse Relation.” In proceedings of International Joint Conference on Natural Language Processing, pp. 45-50.
[159] Sharma S, Srinivas PY, Balabantaray RC. (2015). “Sentiment analysis of code-mix script.” In Computing and Network Communications (CoCoNet), 2015 pp. 530-534. IEEE.
[160] Vyas Y., Gella S., Sharma J., Bali K., Choudhury m. (2014). “POS Tagging of English-Hindi Code Mixed Social Media Content.” In Proceedings of the Con-ference on Empirical Methods in Natural Language Processing (EMNLP), 974-979
[161] Joshi A, Prabhu A, Shrivastava M, Varma V. (2016). “Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text.” In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2482-2491.
[162] Bhargava R., Sharma Y., Sharma S. (2016). "Sentiment Analysis for Mixed Script Indic Sentences", International Conference on Advances in Computing Communications and Informatics ICACCI 2016, pp. 524-529.
[163] Vilares D., Alonso M A., G´omezRodr´ıguez C. (2015). “Sentiment Analysis on Monolingual, Multilingual and Code-Switching Twitter Corpora.” In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 2–8, Association for Computational Linguistics.
[164] Malgaonkar S., Khan, A., Vichare A. (2017). “Mixed bilingual social media analytics: case study: Live Twitter data.” 1407-1412. 10.1109/ICACCI.2017.8126037.
[165] Konate, A. & Du, R. (2018). “Sentiment Analysis of Code-Mixed Bambara-French Social Media Text Using Deep Learning Techniques, Wuhan Univ. J. Nat. Sci. (2018) 23: 237. https://doi.org/10.1007/s11859-018-1316-z
[166] Abbasi, A., Chen, H., & Salem, A. (2008). “Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums.” ACM Transactions on Information Systems, 26(3), [12]. https://doi.org/10.1145/1361684.1361685
[167] Pakray, P., Majumder, G., & Pathak, A. (2018, January). An HMM Based POS Tagger for POS Tagging of Code-Mixed Indian Social Media Text. In Annual Convention of the Computer Society of India (pp. 495-504). Springer, Singapore.
[168] Sane, S. R., Tripathi, S., Sane, K. R., & Mamidi, R. (2019, June). Deep Learning Techniques for Humor Detection in Hindi-English Code-Mixed Tweets. In Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 57-61).
[169] Swami S., Khandelwal A., Singh V., Akhtar S., Shrivastava, M. (2018). “A Corpus of English-Hindi Code-Mixed Tweets for Sarcasm Detection.” eprint arXiv:1805.11869, CICLing 2018
[170] Vijay D., Bohra A., Singh V., Akhtar S., Shrivastava M. (2018). “A Dataset for Detecting Irony in Hindi-English Code-Mixed Social Media Text.” EMSASW@ESWC, volume 2111 of CEUR Workshop Proceedings, page 38-46, CEUR-WS.org, (2018).
[171] Kamble S., Joshi A. (2018). “Hate Speech Detection from Code-mixed Hindi-English Tweets Using Deep Learning Models. arXiv preprint arXiv:1811.05145, 2018.
[172] Guyon, I. and Elisseeff, A. (2006). “An introduction to feature extraction.” In Feature extraction (pp. 1-25). Springer, Berlin, Heidelberg.
[173] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R. and Bengio, Y.

(2015). “Show, attend and tell: Neural image caption generation with visual attention.” In International conference on machine learning (pp. 2048-2057).
[174] Kumar A., Garg G. (2019) Systematic Literature Review on Context-Based Sentiment Analysis in Social Multimedia, Multimedia tools and Applications, https://doi.org/10.1007/s11042-019-7346-5
[175] Kumar A., Jaiswal A. (2019). Systematic literature review of sentiment analysis on Twitter using soft computing techniques, Concurrency Computat Pract Exper. https://doi.org/10.1002/cpe.5107
[176] Kumar A., Sharma A., (2017). Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence. Webology, Vol.14, No. 2.
[177] Kumar A, Jaiswal A. (2017). Image sentiment analysis using convolutional neural network. International Conference on Intelligent Systems Design and Applications, 464-473.
[178] Gajarla V, Gupta A. (2015). Emotion Detection and Sentiment Analysis of Images. Georgia Institute of Technology.
[179] Girshick R, Donahue J, Darrell T, Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 580-587.
[180] Cai G, Xia B. (2015). Convolutional neural networks for multimedia sentiment analysis. In Natural Language Processing and Chinese Computing. 159-167.
[181] Datta R, Joshi D, Li J, Wang JZ. (2006). Studying aesthetics in photographic images using a computational approach. European Conference on Computer Vision, 288-301.
[182] Jia J, Wu S, Wang X, Hu P, Cai L, Tang J. (2012). Can we understand van gogh'smood?: learning to infer affects from images in social networks. Proceedings of the 20th ACM international conference on Multimedia 857-860.
[183] Marchesotti L, Perronnin F, Larlus D, Csurka G. (2011). Assessing the aesthetic quality of photographs using generic image descriptors. IEEE International Conference on Computer Vision, 1784-1791.
[184] Li B, Feng S, Xiong W, Hu W. (2012). Scaring or pleasing: exploit emotional impact of an image. Proceedings of the 20th ACM International Conference on Multimedia, pp. 1365-1366.
[185] Li B, Xiong W, Hu W, Ding X. (2012). Context-aware affective images classification based on bilayer sparse representation. Proceedings of the 20th ACM International Conference on Multimedia, 721-724.
[186] Machajdik J, Hanbury A. (2010). Affective image classification using features inspired by psychology and art theory. Proceedings of the 18th ACM international conference on Multimedia, 83-92)
[187] Lu X, Suryanarayan P, Adams Jr RB, Li J, Newman MG, Wang JZ. (2012). On shape and the computability of emotions. Proceedings of the 20th ACM International Conference on Multimedia, 229-238.
[188] Yanulevskaya V, van Gemert JC, Roth K, Herbold AK, Sebe N, Geusebroek JM. (2008). Emotional valence categorization using holistic image features. In ICIP, 101-104.
[189] Yanulevskaya V, Uijlings J, Bruni E, Sartori A, Zamboni E, Bacci F, Melcher D, Sebe N. (2012). In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. Proceedings of the 20th ACM International Conference on Multimedia, 349-358.
[190] Vonikakis V, Winkler S.(2012). Emotion-based sequence of family photos. Proceedings of the 20th ACM International Conference on Multimedia, 1371-1372.
[191] Kumar A, Sangwan S R. (2019). Rumor Detection Using Machine Learning Techniques on Social Media. International Conference on Innovative Computing and Communications, 213-221.

List of Publications

Journals:

SCI/SCIE:

1. Kumar A., Garg G. **(2019).** *"Systematic Literature Review on Context-Based Sentiment Analysis in Social Multimedia."* Multimedia Tools and Applications. https://doi.org/10.1007/s11042-019-7346-5. Springer**. [SCIE- Impact Factor: 2.101]**

2. Kumar A., Garg G, (**2019**). *"Sentiment Analysis of Multimodal Twitter Data."* Multimedia Tools and Applications. 78, 24103–24119 (2019), https://doi.org/10.1007/s11042-019-7390-1, ISSN: 1380-7501. Springer. **[SCIE- Impact Factor: 2.101]**

3. Kumar A., Garg G, (**2019**). "Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets", Journal of Ambient Intelligence and Humanized Computing,1-16, https://doi.org/10.1007/s12652-019-01419-7, ISSN: 1868-5137. Springer **[SCIE- Impact Factor: 1.910]**

4. Jain DK, Kumar, A., Garg, G. (**2020**). "Sarcasm Detection in Mash-up Language using Soft-Attention based Bi-directional LSTM and feature-rich CNN", Applied Soft Computing, https://doi.org/10.1016/j.asoc.2020.106198, Elsevier ,ISSN: 1568-4946. **[SCIE-Impact Factor: 4.837].**

SCOPUS:

5. Kumar, A., Garg G. "*Fine Grain Sentiment Grading Of User-Generated Big Data Using Contextual Cues"* (**2020**). World Review of Entrepreneurship, Management and Sustainable Development. Inderscience. **[SCOPUS, In Press]**

Conferences

1. Kumar, A., Garg G. (**2019**). Sarc-M: Sarcasm Detection in Typo-graphic Memes (March 14, 2019). International Conference on Advances in Engineering Science Management & Technology (ICAESMT) – 2019. http://dx.doi.org/10.2139/ssrn.3384025. Elsevier [**Presented in March' 2019]**.

2. Kumar, A., & Garg, G. **(2020)**. "*The Multi-faceted Concept of Context in Sentiment Analysis".* In Cognitive Informatics and Soft Computing (pp. 413-421). doi: 10.1007/978-981-15-1451-7_44. Springer, Singapore. **[Presented in April' 2019]**

3. Kumar A., Garg G. (**2020**) Sarcasm Detection Using Feature-Variant Learning Models. In Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering, vol 605. https://doi.org/10.1007/978-3-030-30577-2_61. Springer, Cham. **[Presented in July' 2019].**

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-019-7346-5

Systematic literature review on context-based sentiment analysis in social multimedia

Akshi Kumar [1] · Geetanjali Garg [1]

Received: 23 November 2018 / Revised: 17 January 2019 / Accepted: 5 February 2019
Published online: 23 February 2019

Abstract

The opinion seeking behavior of people for good decision making has greatly enhanced the importance of social media as a platform for exchange of information. This trend has led to a sudden spurt of information overflow on the Web. The huge volume of such information has to be technically processed for segregating the relevant knowledge. Sentiment analysis is the popular method extensively used for this purpose. It is defined as the computational study of mining the opinions from the available content about the entity of interest. Existing Sentiment analysis techniques quite efficiently capture opinions from text written in syntactically correct and explicit language. However, while dealing with the informal data, limitation has been observed in performance of sentiment analysis techniques. With a view to deal with the imperfect and indirect language used by the netizens, it has become necessary to work on improvement in the existing sentiment analysis techniques. In this regard, the conventional sentiment analysis techniques have shown some improvement on applying the appropriate context information. However, still there is ample scope for further research to find the relevant "context" and applying it to a given scenario. This systematic literature review paper intends to explore and analyze the existing work on the context-based sentiment analysis and to report gaps and future directions in the said research area.

Keywords Context · Sentiment analysis · Social multimedia · Systematic literature review

1 Introduction

With the advent of Web 2.0, usage of social media to gather opinions has taken a steep rise. The users across social media have contributed a lot in generating huge volume of voluntarily

✉ Geetanjali Garg

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-019-7390-1

Sentiment analysis of multimodal twitter data

Akshi Kumar[1] · Geetanjali Garg[1]

Received: 8 October 2018 / Revised: 28 January 2019 / Accepted: 19 February 2019
Published online: 12 March 2019

Abstract
Text-driven sentiment analysis has been widely studied in the past decade, on both random and benchmark textual Twitter datasets. Few pertinent studies have also reported visual analysis of images to predict sentiment, but much of the work has analyzed a single modality data, that is either text or image or GIF video. More recently, as the images, memes and GIFs dominate the social feeds; typographic/infographic visual content has become a non-trivial element of social media. This multimodal text combines both text and image defining a novel visual language which needs to be analyzed as it has the potential to modify, confirm or grade the polarity of the sentiment. We propose a multimodal sentiment analysis model to determine the sentiment polarity and score for any incoming tweet, i.e., textual, image or info-graphic and typographic. Image sentiment scoring is done using SentiBank and SentiStrength scoring for Regions with convolution neural network (R-CNN). Text sentiment scoring is done using a novel context-aware hybrid (lexicon and machine learning) technique. Multimodal sentiment scoring is done by separating text from image using an optical character recognizer and then aggregating the independently processed image and text sentiment scores. High performance accuracy of 91.32% is observed for the random multimodal tweet dataset used to evaluate the proposed model. The research further demonstrates that combining both textual and image features outperforms separate models that rely exclusively on either images or text analysis.

Keywords Multimodal text · Twitter · Sentiment · Context-aware · Optical character recognition

1 Introduction

The current affordable and ubiquitous generation of Web provides substantial amount of opinionated social big data which facilitates decision making. Sentiment Analysis has gained

✉ Akshi Kumar
akshikumar@dce.ac.in

Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01419-7

ORIGINAL RESEARCH

Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets

Akshi Kumar[1] · Geetanjali Garg[1]

Received: 21 February 2019 / Accepted: 1 August 2019

Abstract

Sarcastic expressions tend to flip the polarity of posts when being analyzed for sentiments. Detecting sarcastic tone, which conveys a sharp, bitter, or cutting expression, remark or taunt in natural language is tricky even for humans, making its automated detection more arduous. Computational models for sarcasm detection have often relied on the content of utterances in isolation whereas using contextual information definitely improves it. This work is a preliminary to understand the what, how and why of using context in sentiment analysis. The concept of 'context in use' is described by exemplifying content-based local and global context to predict sarcasm in user-generated social textual data. In this research Twitter data of benchmark SemEval 2015 Task 11 and nearly 20 k posts from Reddit are classified as sarcastic or non-sarcastic using three predictive learning models. The first model is based on the conventional Term Frequency-Inverse Document Frequency (TF-IDF) weighting which is trained over three classifiers, namely the Multinomial Naïve Bayes, Gradient Boosting and Random Forest and Ensemble Voting is utilized to generate the output. In the second model, a combination of semantic (sentiment) and pragmatic (punctuation) features are considered to model the context along with the top-200 TF-IDF features and results are observed using five baseline classifiers (Decision Tree, Support Vector Machine, Random Forest, K-Nearest Neighbour and Multi Layer Perceptron). The final model uses deep learning, that is, the Long–Short-Term-Memory (LSTM) and its variant Bi-directional LSTM applying GloVe (Global Vectors for Word Representation) for building semantic word embeddings and learning context. The empirical study using training set and test set performance metrics (Accuracy, Recall, Precision, F1 Score) is done to compare the three learning models for sarcasm classification within two datasets, and it is observed that Bi-directional LSTM model has the highest accuracy of 86.32% and 82.91% for the Twitter and Reddit datasets respectively.

Keywords Sarcasm · Shallow classifier · Deep classifier · Social media

1 Introduction

The pervasive social web is a rich source of real-time, opinionated user-generated multimodal content available online on resources like review forums and microblogging sites. Automated text analytics using this opinion-rich data has many promising practical applications, such as sentiment analysis (Pang and Lee 2008; Pak and Paroubek 2010; Liu 2015), aggression detection (Chen et al. 2018), and rumour detection (Zubiaga et al. 2017). Sentiment Analysis has been thriving to facilitate knowledge extraction for decision making within this social setting. It is often described as typical text classification task that indispensably relies on the understanding of the human language and emotions expressed via textual or non-textual content. But human expressions are extremely complicated as statements can convey a wide range of emotions, and requires contextualization of words to bridge the gap between what you've experienced and what you are trying to say. Accurately determining polarity from the legit literal text is imperative for real-time sentiment analysis. However, the growing use of emblematic language markers such as punctuations (awesome!!!!!!), emojis (,, ❤),wordplay (greattttt for great), creative spellings (2 day for today), Internet slangs (OMG for 'Oh My God') (Satapathy et al. 2017), code-switching (Nilep 2006) and code-mixing (Barman et al. 2014; Das and Gambäck 2014) are some commonly seen phenomenon which increase the

✉ Akshi Kumar
akshikumar@dce.ac.in

Applied Soft Computing Journal 91 (2020) 106198

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN

Deepak Jain[a], Akshi Kumar[b,*], Geetanjali Garg[b]

[a] Key Laboratory of Intelligent Air-Ground Cooperative Control for Universities in Chongqing, College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China
[b] Department of Computer Science & Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Article history:
Received 15 October 2019
Received in revised form 24 January 2020
Accepted 21 February 2020
Available online 6 March 2020

Keywords:
Sarcasm
Code-switch
Mash-up
Social media
Deep learning

ABSTRACT

Analyzing explicit and clear sentiment is challenging owing to the growing use of emblematic and multilingual language constructs. This research proposes sarcasm detection using deep learning in code-switch tweets, specifically the mash-up of English with Indian native language, Hindi. The proposed model is a hybrid of bidirectional long short-term memory with a *softmax* attention layer and convolution neural network for real-time sarcasm detection. To evaluate the performance of the proposed model, real-time mash-up tweets are extracted on the trending political (#government) and entertainment (#cricket, #bollywood) posts on Twitter. The randomly sampled dataset contains 3000 sarcastic and 3000 non-sarcastic bilingual Hinglish (Hindi + English) tweets. Feature engineering is done using pre-trained GloVe word embeddings to extract English semantic context vector, hand-crafted features using subjective lexicon Hindi-SentiWordNet to generate the SentiHindi feature vector and an auxiliary pragmatic feature vector depicting the count of pragmatic markers in tweet. Performance analysis is done to compare and validate the proposed ^{softAtt}BiLSTM-$^{feature\text{-}rich}$CNN model. The model outperforms the baseline deep learning models with a superior classification accuracy of 92.71% and F-measure of 89.05%.

1. Introduction

The number of social network users worldwide continue to grow and is expected to pass 3 billion in 2020.[1] The social, interactive computer-mediated technologies, such as, Twitter, Tumblr, Google+, Facebook, Instagram, Snapchat, YouTube etc. which enable users to create, post and share all types of multimedia text are gaining users globally. Social media serves as a communication channel and a social listening, awareness, activism & feedback tool for stakeholder engagement and cooperation. Keeping tabs on social media sentiment, that is, *'the online mood'* is a key part of social media listening. Sentiment analysis [1,2] is a key component of the social listening tool for measuring and reporting on the tone or sentiment of your social mention. It is the use of natural language processing (NLP) to analyze social conversations online and determine deeper context as they apply to a mention (topic, brand or theme). Determining the literal exactitude and opinion polarity in text is imperative for real-time sentiment analysis. Sentiment analysis task attains fitting results with literal language where articulation is same as anticipated interpretation. The non-standard vocabulary with informal textual content and noise currently define the content strategy of social media [3–5]. The common use of emblematic language markers such as punctuations (great!!!!!!), emojis (😌, 👎, ❤),micro-text [6] which includes wordplay (*suppppermrrr* for *super*), neologisms (*gr8* for *great*), Internet slangs (*FYI* for *for your information*), code-switching [7] and code-mixing [8] makes automated text analytics computational intensive.

As one of the most popular literary genre on social media, comedy defines the use of comic literary devices, which include wit, fun, humor, sarcasm, satire, pun, irony and non-sense. We characterize this as the 'Comedy Cube' (Fig. 1).

Satire is a literary genre that relies heavily on irony, wit, and sometimes sarcasm. *Irony* describes the difference between what is expected and what actually happens and Wit is about understanding and intelligence. *Sarcasm* a sharply ironical taunt; sneering or cutting remark. It is the use of wit or ridicule to taunt, mock or contemptuously strike at a target. Irony is a common tool for sarcasm, but it is by no means the only one. Often, a mere laugh can be sarcastic, without any use of words. With reference to sentiment analysis, "Sarcasm is defined as a specific

The Multifaceted Concept of Context in Sentiment Analysis

Akshi Kumar and Geetanjali Garg

Abstract The contemporary web is about communication, collaboration, participation, and sharing. Currently, the sharing of content on the web ranges from sharing of ideas and information which includes text, photos, videos, audios, and memes to even gifs. Moreover, the language and linguistic tone of user-generated content are informal and indistinct. Analyzing explicit and clear sentiment is challenging owing to language constructs which may intensify or flip the polarity within the posts. Context-based sentiment analysis is the domain of study which deals with comprehending cues which can enhance the prediction accuracy of the generic sentiment analysis as well as facilitate fine grain analysis of varied linguistic constructs such as sarcasm, humor, or irony. This work is preliminary to understand the what, how and why of using the context in sentiment analysis. The concept of 'context in use' is described by exemplifying the types of context. A strength–weakness–opportunity–threat (SWOT) matrix is made to demonstrate the effectiveness of context-based sentiment analysis.

Keywords Sentiment analysis · Context · Social media · SWOT

1 Introduction

Undeniably, the cross-platform, cross-lingual, multimodal social web is omnipresent. Popular sites such as Twitter, Facebook, and Instagram have become a critical part of our daily lives to share content, stay connected, and gain insights. The expansive user base on such social networking sites generates voluminous data which can be intelligently filtered and analyzed for building a real-time knowledge discovery framework. The user-generated data can be used by organizations to analyze the trends of market, opinion for the elections, recommendation of the products, and

A. Kumar · G. Garg (✉)
Department of Computer Science and Engineering, Delhi Technological University, Delhi, India
e-mail: geetanjali.garg@dtu.ac.in

A. Kumar
e-mail: akshikumar@dce.ac.in

Sarcasm Detection Using Feature-Variant Learning Models

Akshi Kumar and Geetanjali Garg(✉)

Department of Computer Science and Engineering,
Delhi Technological University, Delhi, India
akshikumar@dce.ac.in, geetanjali.garg@dtu.ac.in

Abstract. Sentiment Analysis is the text classification tool that analyses a sentiment, message, emotion, attitude and tells whether the sentiment is positive, negative or neutral. The prime challenging aspect of sentiment analysis is the presence of sarcasm in message. Sarcasm is one kind of sentiment that is expressed verbally through the use of rolling of eyes and tonal stress. It consist of words mean the opposite of what user want to convey in order to be funny, or to show some irritation. The active online users and their reviews on websites are large in number so it is hard to detect even for humans, so in order to achieve error-free sentiment analysis it is imperative for machines to detect it accurately. The paper proposes the use of three different classes of features to help computers identify sarcasm reasonably well. In this paper, we intend to implement and empirically analyze number of computing techniques like Support Vector Machine, Decision Trees, Logistic Regression, Random Forest, K-Nearest Neighbors and Neural Networks for sarcasm detection on social media. The experimentation was done using three datasets i.e. SemEval 2015 Twitter benchmark dataset; random tweets collected using the Streaming API and a publicly available dataset of Reddit posts. The datasets provide interesting insights into how different forms of social media use the tool of sarcasm differently. The evaluated results were based on the performance measures like precision, recall, accuracy and F score. Amongst all, Twitter datasets had achieved the highest accuracy of around 91% to 92%, while the Reddit dataset had obtained peak accuracy of 80%.

Keywords: Sarcasm · Machine learning · Lexicons · Twitter · Reddit

1 Introduction

The advent of social media has revolutionized how people interact amongst themselves and has presented a new paradigm for businesses to engage with their customers [1]. The vast volume of exchange of ideas and thoughts over these platforms has made it necessary to evaluate how online trends transcend to real world changes. Hence, there has been a great interest on social media platforms in recent times to develop models for sentiment analysis [2–8], humor detection, and evaluation of consumer engage-

International Conference on Advanced Engineering, Science, Management and Technology – 2019 (ICAESMT19)

Sarc-M: Sarcasm Detection in Typo-graphic Memes

Akshi Kumar[a], Geetanjali Garg[b]*

[a,b]*Dept. of Computer Science & Engineering, Delhi Technological University, Delhi, India-110042*
**akshikumar@dce.ac.in*

ARTICLEINFO

Article history:

Keywords:
Sarcasm
Memes
OCR
Sentiment

ABSTRACT

Detecting sarcastic tone, which conveys a sharp, bitter, or cutting expression, remark or taunt in natural language is tricky even for humans, making its automated detection more arduous. The growing use of typo-graphic images, that is text represented as an image further characterizes the power of expressiveness in online social data. This research proffers a model Sarc-M, a sarcastic meme predictor, for sarcasm detection in typo-graphic memes using supervised learning based on lexical, pragmatic and semantic features. The learning model is evaluated using five different classifiers and the results are evaluated using a balanced dataset of typo-graphic images, called MemeBank, scrapped from Instagram. The contribution of the research is two-fold, firstly, typo-graphic text is extracted using optical character recognizer and then analyzed for sarcasm and secondly for detecting sarcasm the need of contextual information is explored, that is, contextual cues such as frequency of punctuations and sentiment words are considered as features. The best sarcasm prediction model for typo-graphic memes is built using Multi Layer perceptron which achieves an accuracy of approximately 88%.

Peer review under responsibility of Uttaranchal University, Dehradun.

1. Introduction

More recently, as memes and GIFs dominate the social feeds; typo-graphic visual content has become a considerable element of social media. 'Meme' is a viral image or video often altered by internet users for humorous effect. These convey human expressions but with a wide range of emotions, and often require context to fully understand humor and sarcasm. "Sarcasm is a type of sentiment where people express their negative feelings using positive or intensified positive words in the text" [Bharti et al., 2016]. It is an expression representative of conflict between the apparent and the applied. Memes are topic-dependent and highly contextual, therefore, polarity shift and other contextual clues can help detect sarcasm from text and improve the generic sentiment classification of typo-graphic social data. For example, as in the post shown in Fig. 1, "Being nice" demonstrates a conflict between the obvious state of "hardest part of the job". This inconsistency, contrast and shifts within the polarities of sentiments validate sarcasm as a distinctive case of sentiment analysis.

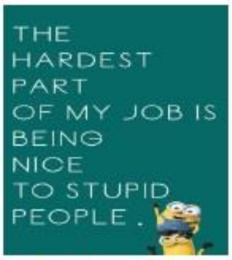


Fig. 1- Example of sarcastic meme

Detecting sarcastic tone is very difficult to accomplish without having a sufficient knowledge of the 'context' of the situation, the specific topic, and the environment [Kumar & Garg, 2019]. Textual sentiment analysis has been widely studied [Pang & Lee, 2008; Kumar & Jaiswal, 2019; Kumar & Sharma,