CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Context Driven Aspect based Citation Sentiment Analysis

by

Muhammad Touseef Ikram

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Computing
Department of Computer Science

2021

Context Driven Aspect based Citation Sentiment Analysis

By

Muhammad Touseef Ikram

(PC141002)

Dr. Denis Helic, Associate Professor
Graz University of Technology, Austria
(Foreign Evaluator 1)

Dr. Hermann Maurer, Professor
Graz University of Technology, Austria
(Foreign Evaluator 2)

Dr. Nayyer Masood
(Thesis Supervisor)

Dr. Nayyer Masood
(Head, Department of Computer Science)

Dr. Muhammad Abdul Qadir
(Dean, Faculty of Computing)

DEPARTMENT OF COMPUTER SCIENCE
CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
ISLAMABAD
2021

Dedicated to my beloved Father Professor Muhammad Ikram

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD
Expressway, Kahuta Road, Zone-V, Islamabad
Phone:+92-51-111-555-666 Fax: +92-51-4486705
Email: info@cust.edu.pk Website: https://www.cust.edu.pk

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in the thesis, entitled "**Context Driven Aspect based Citation Sentiment Analysis**" was conducted under the supervision of **Dr. Nayyer Masood**. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Computer Science, Capital University of Science and Technology** in partial fulfillment of the requirements for the degree of Doctor in Philosophy in the field of **Computer Science**. The open defence of the thesis was conducted on **August 30, 2021**.

Student Name : Muhammad Touseef Ikram (PC-141002) ______________________

The Examining Committee unanimously agrees to award PhD degree in the mentioned field.

Examination Committee :

(a)	External Examiner 1:	Dr. Zahid Halim, Associate Professor GIKI, Topi, Swabi	______________________
(b)	External Examiner 2:	Dr. Hammad Majeed, Associate Professor FAST-NUCES, Islamabad	______________________
(c)	Internal Examiner :	Dr. Abdul Basit Siddiqui Associate Professor CUST, Islamabad	______________________

Supervisor Name :	Dr. Nayyer Masood Professor CUST, Islamabad	______________________
Name of HoD :	Dr. Nayyer Masood Professor CUST, Islamabad	______________________
Name of Dean :	Dr. Muhammad Abdul Qadir Professor CUST, Islamabad	______________________

Author's Declaration

I, **Muhammad Touseef Ikram** hereby state that my PhD thesis titled "**Context Driven Aspect based Citation Sentiment Analysis**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my PhD Degree.

(Muhammad Touseef Ikram)

Registration No: PC141002

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Context Driven Aspect based Citation Sentiment Analysis**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD Degree, the University reserves the right to withdraw/revoke my PhD degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Muhammad Touseef Ikram)

Registration No: PC141002

List of Publications

It is certified that following publication(s) have been made out of the research work that has been carried out for this thesis:-

1. **Ikram, M. T.**, and Afzal, M. T. (2019). "Aspect based citation sentiment analysis using linguistic patterns for better comprehension of scientific knowledge". *Scientometrics*, 119(1), 73-95.

2. **Ikram, M. T.**, Afzal, M. T., and Butt, N. A. (2018). "Automated citation sentiment analysis using high order n-grams: a preliminary investigation". *Turkish Journal of Electrical Engineering Computer Sciences*, 26(4), 1922-1932.

3. **Ikram, M. T.**, Butt, N. A., and Afzal, M. T. (2016). Open source software adoption evaluation through feature level sentiment analysis using Twitter data. *Turkish Journal of Electrical Engineering Computer Sciences*, 24(5), 4481-4496.

Muhammad Touseef Ikram
(PC141002)

Acknowledgement

First and foremost, I want to thank Almighty Allah for His countless blessings upon us. Without His guidance and help, I could not do the research presented in this dissertation. I would also like to thank my supervisor, Dr. Muhammad Tanvir Afzal, for providing me the opportunity to work on the fantastic topic of Citation Sentiment Analysis and guiding me throughout my Ph.D. studies with great patience. I appreciate his effort, time and ideas to make my Ph.D. experience productive. I also express my sincere gratitude towards all those professors who contributed to my academic development at Capital University of Science and Technology (CUST).

I am indebted to all the internal, external and foreign examiners for taking time to read and evaluate my synopsis and thesis, for identifying the shortcomings, and for providing their learned opinion on how to improve upon these.

There are numerous other people who helped me at (CUST) in pursuit of my PhD in one way or the other including the faculty members at Department of Computer Science, managerial, support staff, and the librarian to mention a few. Thank you all.

I owe a lot to my close friends, Naveed Anwer Butt, Tauseef-ur-Rehman and Muhammad Naseer, and my brother Dr. Touqeer Ahmad for being constant source of inspiration and motivation throughout this time. There are so many other well-wishers including friends, colleagues and relatives who remembered me in their prayers. Allah bless you all.

In the end I must mention the being who cares more about me than I do for myself, my mother. I always felt her prayers by my side in the time of despair and frustration, and this feeling gave me energy to put myself together. And last but not the least, I have profound gratitude for my beloved wife Mrs. Sadia Touseef who had to face double dose of educational stresses: from my children Fatima Madni and Ali Hassan as well as from me. May Allah reward you for your sacrifice and selflessness.

Abstract

Scientific papers hold an association with the previous research contributions in the form of citations. The nature of the cited material could be positive, negative, or objective. In this thesis a technique is proposed for the identification of citing author's sentiment towards cited paper by extracting unigram, bigram, trigram and pentagram adjective and adverb patterns from the citation text. After doing part of speech tagging on citation text, I used the sentence parser for the extraction of linguistic features comprising of adjectives, adverbs, and n-grams from the citation text. A sentiment score is then assigned to distinguish them as positive, negative and neutral. In addition, the proposed technique is compared with the manually classified citation text and two commercial tools, namely SEMANTRIA and THEYSAY. The analysis of the results depicted that the proposed approach has achieved comparable results with the commercial counterparts with an average precision, recall and accuracy of 90%, 81.82%, and 85.91% respectively. Further, this thesis presents a novel approach to identify aspect level sentiments. The approach is comprised of two levels. At first level, it extracts the aspects from the citation sentences using the pattern of opinionated phrases around the aspect. At the second level, it detects the sentiment polarity of the identified aspect considering nearby words and associate it with the corresponding aspect category using linguistic rule based approach. The approach consider 'N-gram after', 'N-gram before' and 'N-gram around' features. The results revealed that n-gram around feature performed better than others. It further indicates that SVM outperformed other classifiers for all n-gram models with an average precision 0.82, recall 0.807 and accuracy of 0.89. This thesis also investigates how the citation text and sentiments associated with them are distributed along the IMRaD structure. The analysis of the results depicts that expression of the positive sentiment towards the cited paper is most common at the start of the research paper i.e., "Introduction" followed by the "Discussion" section. The most significant result is that the "Discussion" section is designated with the largest number of negative citation contexts as compared to "Results" and "Introduction" along with majority of objective citation mentions found in "Literature" section.

Contents

List of Figures

List of Tables

Abbreviations

BOW	Bag of Words
CNN	Convolutional Neural Networks
CRF	Conditional Random Fields
IG	Information Gain
IMRaD	Introduction, Methods, Results and Discussion
mRMR	Minimum Redundancy and Maximum Relevancy
RNN	Recurrent Neural Network
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine

Chapter 1

Introduction

1.1 Overview

The number of research publications is exponentially mounting up each passing year. The information management of this scientific literature is a prime focus. Due to the availability of this rich plethora of scientific contributions, it becomes difficult for the researchers and students to be abreast with the opinions expressed in the scientific literature and more explicitly with the citations. Therefore, an important element of this scholarly big data analysis is to keep a track of citations. Extensive work has been carried out for the sentiment analysis of product reviews, movie reviews, narrative text, blog posts, forums, feedback, recommendations, ratings, and comments, etc., [1–3] whereas less emphasis has been placed on the agenda of gaining insight into extracting opinions from the citations [4]. The current research repositories don't provide any capability for summarizing the research citations based on sentiments expressed in them [3]. For instance, before citing a paper a researcher might wish to know what other researchers and the scientific community are saying about that specific scientific contribution. Thus, it ascertains a need for the sentiment analysis system that can analyse the bulk of such citations and provide meaningful information to the researchers and students. Identifying why people cite a particular paper has been a matter of interest and investigation by diversified domains of information sciences, discourse analysis and social sciences. The existing citation analysis techniques bibliometrics, h-index, g-index, a-index, impact factor, etc., are prone to some inherent limitations like they ignore the intention of citing a particular research paper and much of them are quantitative [3, 5, 6]. Because of these reasons they are unable to access the true impact and author's stance and opinion towards the contents of the earlier work. So, there is an ultimate need for citation sentiment/content analysis based on text mining and lexical analysis techniques. The citation text contains the precise and concise analysis of a paper because of the space limitations in papers and due to the high quality of paper in terms of its correctness [7]. Existing approaches on citation sentiment detection rely only on the citation text. But there are certain inherent complexities associated with the sentiment detection from the citation text such as (1) hidden sentiment (2) impartiality of the citation text (3) contrastive expression of the negative sentiment (4) variation in the lexical terms used to express the sentiment in scientific text (5) use of the technical terms and (6) the scope of citation text and (7) being more selective in vocabulary. Another difference between the general sentiment analysis and the citation sentiment detection is because of the singular characteristics of the citations [4, 8]. All these reasons make citation sentiment analysis one of the most complex and challenging fields for sentiment analytics to be applied [1, 4, 9, 10]. There are different aspects of citation analysis research. Citation sentiment analysis is a new and interesting domain that urges investigation of the opinion of the citing author towards the cited text [11, 12]. It first extracts the opinion of citing paper towards the target i.e., cited paper and classifies its sentiment polarity as positive, negative or neutral [1, 4, 5, 9, 13, 14]. Apart from just the citation sentiment classification, there are numerous studies for the citation function classification as 'Influential/Non-Influential', 'Functional/Perfunctory', and 'Contrast/Conflict' [4, 7, 15–19]. One dimension is the summarization of the scientific contribution based on the citations made. Similarly, another dimension is devising the qualitative measures for evaluating the scientific contributions as compared to the quantitative measures. This will lead to developing new applications in bibliometrics and bibliographic research. Further helping in evaluating the impact and influence of journals, scientific articles and individuals by extracting sentiments expressed in citations.

1.2 Research Motivation

The process of finding out the relevance and frequency of citations from the different literary materials, including journals, research papers, articles and books is known as the citation analysis. Generally, the importance or impact of a research paper is usually calculated by calculating the number of citations it receives over the period. But this is mainly the quantitative assessment of the literary work and can be biased because of the Guest Citation or Random Citation. For the proper evaluation and assessment of the paper, the cited text is of utmost importance as the researchers excerpt the general ideas and concepts from different papers and mention the outcome and findings of the paper in their discussion. But this does not mean that the referring paper would always be describing the paper with positive words. Sometimes if the referring paper has achieved some comparable results as of the referred paper, then the authors would not give much positive response to the previous one. Therefore, to address this issue sentiment analysis of the citation text is necessary for rating the citations based on their polarity level. In the existing literature different single machine learning techniques have been applied to determine the citation sentiment e.g., Support Vector Machine [1, 4, 5, 9, 17–19], Naïve Bayesian [7, 13, 15], Maximum Entropy classifier [16], Logistic Regression [17, 18]. Different features like n-grams [1, 5, 7, 9, 16, 19], dependency relations, sentence splitting, negation features, physical features (location, popularity, density) [4, 15–18], contextual polarity features, etc. are widely used in literature. But as we know "two are better than one", similarly hybrid techniques are better than the single algorithm. All the above-stated techniques have their own associated advantages and disadvantages. Therefore, one aspect of our proposed research work is design some hybrid methodology by using machine learning and data mining techniques for citation sentiment classification. Further, it is established in the literature that higher-order n-grams i.e., longer phrases tend to be less ambiguous in terms of their polarity because they capture the short distance negations and positional context. Higher-order n-grams as features can achieve comparable or improved classification accuracy than state of the art on large scale datasets. This method can be further explicated using the following example. Considering the example of a citation sentence “The proposed approach yields better accuracy”. Its unigram: ’The’, ’proposed’, ’approach’, ’yields’, ’better’, ’accuracy’ in which a single word is considered. Its bigram: ’The proposed’, ‘proposed approach’, ‘approach yields’, ‘yields better’, ‘better accuracy’ in which word pairs are considered. Its trigram: ’The proposed approach’, ’proposed approach yields’, ‘approach yields better’, ‘yields better accuracy’ where a sequence of words having count equal to 3 is considered. Higher-order n-grams refer to 4-gram, 5-gram and so on. Thus, analysing the research outcome of several authors in this thesis I intend to extend the citation sentiment classification using unigram, bigram, trigram and pentagram adjective and adverb and their combinations. Given a research paper with a long list of citations, the model could identify the most influential aspects and generate the aspect-based sentiment summary of the research paper. Consider the following citation sentence example: ‘The technique is efficient but the dataset is not very comprehensive’. Here the citing author expresses the conflicting sentiment about the two aspects of research – aspect technique connotes positive opinion whereas second aspect dataset is referred to as negative. This example depicts the importance of fine-grained sentiments, in that they depict a citing authors preferences while citing the research paper that drives the linkage between two papers. The aim is the extraction of all possible and relevant aspects from the citation sentences and then grouping the synonyms. For example, “Technique is efficient” and “Approach yields the better result”, extract the words “technique” and “approach” which represent a cited research’s aspect, then group them into one category as they both point to the same thing.

The next task is identifying the corresponding sentiment for each identified and extracted aspect. This will facilitate in measuring the aspect-based sentiment strength or intensity of citations (the level of positive and negative citations and aspects). The summarization not only intends to abridge the core idea of a cited

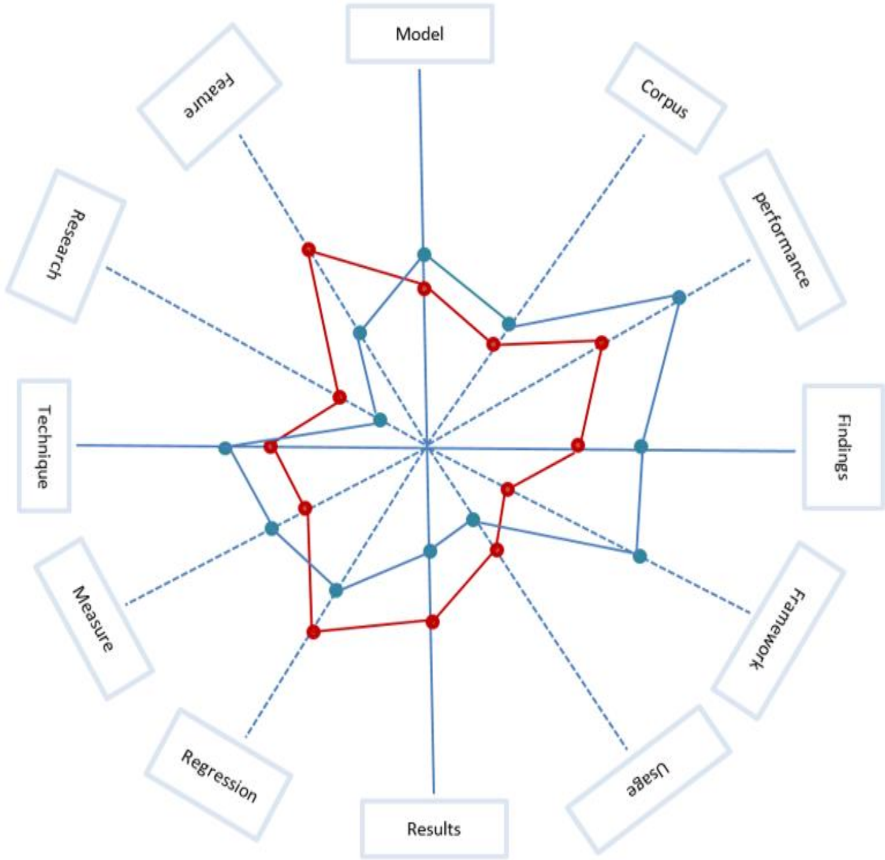

FIGURE 1.1: Example: Aspect based Citation Sentiment Profile.

research paper but also evaluates the specific facets/aspects of the research (e.g., technique, dataset, results, algorithm, corpus etc.). The motivation is to identify the fake opinions towards the cited research. The example presented in figure 1.1 describes the aspect-based sentiment summary of a research paper. This will help in better understanding the research paper along with its cited aspects.

There are only a few studies concentrating on strength of the sentiment by identifying the adjectives and ranking the citations based on the extracted citation polarity. Another motivation of the proposed research is to develop some automated approach in identifying the citing author's sentiment towards the cited literature expressed in the citation sentence using different machine learning techniques. Furthermore, I want to identify the polarity of the citation when it is used in a particular context. I also intend to conduct a comparative experimental study

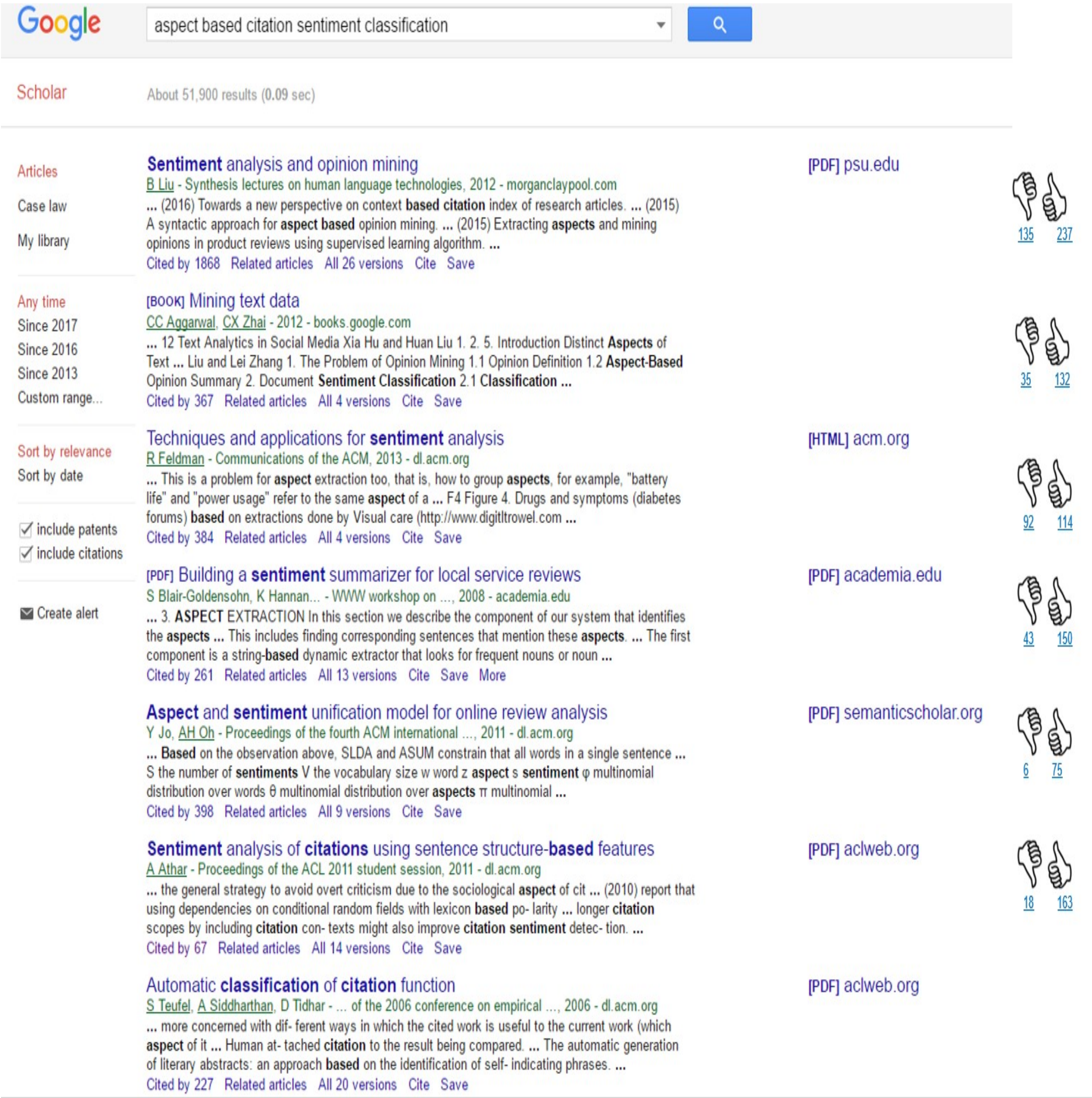

Figure 1.2: Research Motivation.

across multiple domains of corpus for validation of the consistency of our findings.

1.3 Critical Analysis and Problem Formulation

In this research work, I addressed three important problems in the field of Citation Sentiment Analysis. In the following headings, critical analysis has been presented which have led me to work in these two research dimensions. This helps in better understanding the problem. This critical analysis becomes a pathway to deep dive into research domain which is critical for analysis and understanding.

1.3.1 Automated Citation Sentiment Analysis using high order n-grams

In the domain of natural language processing and probability, n-grams are widely used in features for opinion mining. N-grams are the contiguous and continuous sequence of n-items, terms or objects from a given document, sequence of text or speech, which is used as a feature to get sentiment cues from the text [2, 20, 21]. N-grams can be of fixed or variable sequence either at the character or token level. There are several advantages of using n-grams: (1) using n-gram as a basic term ensures fault tolerance in spelling mistakes and require no prior knowledge; (2) using n-grams the system can achieve language independence; (3) dictionaries, grammars and regulations become unnecessary; (4) systems considering n-grams become free from stop words; and (5) there is no need to separate the characters from the words. The main limitation associated with the utilization of n-grams for sentiment classification is that semantic information is partially lost, and it cannot handle the negations effectively [7, 22]. The n-grams of sizes two and three are most commonly used in the state-of-the-art literature [4, 5, 20]. N-grams of size 1 or 2 provide better results when combined with some other features but they fail in some of the cases. For example, if we analyse the following citation text with the unigram approach. "The model is not applicable" This would result in the neutral polarity of the citation text because of the presence of one positive polarity word i.e., "applicable" and one negative polarity word i.e., "not". Similarly, when we analyse the above text using the bigram approach, this will classify it as negative because of the presence of the word "not applicable" which is the correct classification. Considering the example "This approach is not very impactful but produce excellent word alignments", highlights negation, intensification and contrast about the research contribution. So, this scenario depicts that when we consider using the higher-order n-grams, the results are expected to be better. The nuance of using higher-order n-grams is that they not only consider the marked word but capture the short term positional information which helps achieve the state of the art results. I hypothesize, that the use of the high order n-grams might help detect the citation sentiment. Secondly, it is stated in the literature that technical terms play a dominant role in the scientific citation text [1]. Sometimes, sentiment is associated with these terms. For example, if we consider the citation text "the state of the art sentiment classification systems use n-gram features...". In such a situation considering only the lower-order n-grams will not be suitable for classifying the citation sentence because shorter phrases tend to be ambiguous in terms of their polarity. Therefore, in this study, I attempt to classify the citation sentences using higher-order n-grams and examine their impact on citation sentiment detection. Higher-order n-grams are more precise and deterministic as compared to lower-order n-grams. Using long fragments might help capture the polarity information as there is a direct relationship between the fragment-length and the accuracy of the sentiment classification. For the experimental setup, I have set the maximum fragment length to 5. In the experiments with bigram, trigram and pentagram features and their combinations, I also incorporate lower-order n-grams (unigrams and bigrams). The most specific reason for doing so is to demonstrate the prediction accuracy of using higher-order n-grams on large scale sentiment analysis problems. This thesis implicate the identification of the citing author's sentiment towards the cited paper by extracting unigram, bigram, trigram and pentagram adjective and adverb patterns from the citation text. After POS tagging the citation text, I used the sentence parser for the extraction of linguistic features comprising of adjectives, adverbs, and n-grams from the citation text. In this process, irrelevant information e.g., special characters, HTML tags, spelling mistakes etc., is filtered out. At this step this extraneous information is removed. The next step is the removal of stop words from citation sentences.The next function that is performed on the corpus is stemming in which the words are reduced to their term/root form. I split the citation text into tokens of high order n-grams, i.e., n=5 from left to right. Then I find the sentiment score of these terms based on the SentiWordNet 3.0 dictionary which is later classified as positive, negative and neutral based on the sentiment score.This is helpful in determining the right sentiment orientation. Sentiment strenght is also necessry for the citation analysis. Aspect detection can also be helpful performing analysis.

1.3.2 Determining the impact of Citation Contextual Clues in Citation Sentiment Detection

The literature review reveals, that the writing style of the scientific article is objective which results in classifying most of the citations as objective or neutral [4, 23]. Most of the authors assert to be objective or generic in writing reviews towards the cited article. They usually hide the negative sentiment as duplicitous praise which makes sentiment detection even more perplexing [1, 4, 5, 24, 25]. In such a scenario, considering only the citation sentence will not be enough. To address this problem, I take into account the citation context as an extended scope for citation sentiment detection. Citation context refers to on-topic, text or sentences which surround the citation text or references in the scientific articles. The text which surrounds the citation sentence is considered to be its local context. The citation context with a varied window size ranging from a few words to several sentences centred-around the citation sentences have been used previously [19]. The existing approaches to the citation context extraction can be classified into two groups. Symmetric window approach which considers a window of words i.e., n tokens before and after the citation sentence. The second approach is the sentence-based approach where the context is considered based on the n sentences before and after the citation sentence. As per the existing literature, citation context consisting of window of 3 sentences is considered sufficient for classification decision. Most of the existing works on the citation sentiment detection consider only the citation sentence and merely ignore its local context which might undermine its classification accuracy. I propose that considering citation context will help in improving the BOW (bag-of-words) model and unveil the hidden sentiments regarding the cited work. For example, considering the citation sentence "[] has recently proposed a simpler SVM-based algorithm for analogical classification called Pair-Class" and ignoring its subsequent anaphora will not indicate a clear sentiment regarding the cited work. The posterior citation text "it does not adopt a set-based or distributional model of relational similarity" bears a negative sentiment orientation. This ascertains that the sentiment associated with a citation sentence is not static but it is dependent on the particular context in which it has been cited [26, 27]. Summing up, I can say that sentiment of a citation text can be conveyed via both of its context and conceptual semantics. Therefore, I hypothesize that the text before and after the citation text may have an impact on the subjectivity of the citation text. Further, I want to investigate the correlation between the posterior context dependent citation sentences and prior context independent sentences. I also study the impact of different sizes of contextual windows on a generic dataset for identifying the citation sentiment. I intend to determine the optimal contextual window for classifying the implicit citations which don't constitute any explicit anchor to the target paper. In this study, I further examine the distribution of citation sentiment along the progression of different sections of the research paper and identify low and high sentiment density zones. By making an investigation of sentiment context and its location, we can not only understand the intention why citing authors cite in such a way but also ascertain the pattern how these sentiments are distributed across IMRaD. The motivation in this regard is to elucidate the existence of a relation between the distribution of the citation sentiment and the argumentative structure of the research article.

1.3.3 Sentiment Associations based on Citation Aspects

Every research study has some sort of linkage with other contemporary studies in the literature. Researchers cite state-of-the-art studies to acknowledge the research contribution, which establishes a relationship between citing and cited papers [28, 29]. Such connections are normally termed as citation mention, reason, purpose or function and are well studied by many researchers from different viewpoints in the past [1, 4, 30, 31]. The citation text is valuable and of utmost importance for the qualitative assessment of the paper, but its sheer and thriving size make information discovery a challenging task. One of the main directions of citation function classification is citation sentiment analysis wherein citation-based learning analytics have been adopted to classify citations [23, 32]. Most of the previous studies concurred on an oversimplified assumption that if paper A

Table 1.1: Example of Aspects from Citation Sentence

S.No	Citation Sentence	Reference	Aspects
1	"Our similarity method is similar, but simpler, to that used by, which report very good results on similarity datasets."	(Hughes and Ramage,2007)	Method
2	"The IBM models 1-5 produce excellent word alignments with increasing algorithmic complexity and some performance issues."	(Brown et al., 1993)	Performance, Word Alignments

cites paper B, it expresses a sentiment or opinion towards the cited work at the document or sentence level [4, 33]. However, merely applying the document-level sentiment analysis is not sufficient to be unequivocally beneficial for evaluating a research contribution. This is because citation sentences are less often just positive or negative in their sentiment orientation as a whole [14, 34]. Instead, these usually focus on an aspect of the cited work, which expounds that each citation sentence holds some information pertaining to the positive or negative aspect of a cited work [35, 36]. Therefore, to be able to leverage the available information, there is an extensive need of applying learning analytics and cognitive computing for summarizing and processing the scholarly big data based on certain aspects of the cited work [17, 37]. The identification of sentiment polarity at a fine-grained aspect or feature of the cited work is still an unascertained research area, which can open up new avenues in bibliometric and bibliographic research [38, 39]. Various aspects of the cited work like technique, corpus, method, task, concept, measure, model, tool, performance can be alluded by the citation function along with its sentiment orientation mentioned as positive, negative or objective [35, 40]. Extending the sentiment analysis to aspect level can assist the researchers to identify 'material' aspects of the cited work and could be a potential performance indicator for advanced decision-making capabilities with reference to research adoption and penetration [41, 42]. Few examples from the citation corpus to highlight the aspect detection are presented in Table 1.1:

For instance, from the citation sentence snippet presented in Table 1.1, one can easily infer that just performing the sentence level sentiment classification cannot depict the pellucid aspects of cited work. In the first example, the citing author has appraisal sentiment towards the "methodology" and "results" of the cited work, whereas in the second example, the author has encouraging sentiment for the generated "alignments" but at the same time, he is criticizing the "complexity" and "performance" of the model. Such sort of situation demands a coherent way to perform aspect level citation sentiment classification to discover the hidden relationship between different features of the cited research and its expressed opinion. To attain sufficient insights into the citations and determine the qualitative impact of the research paper, citation text summaries should include the sentiment information on the aspect level, as compared to just one overall sentiment score. This thesis presents a novel approach to tackle the aspect level sentiments about cited papers by harnessing the citation text from citing papers. For this purpose, I extract different aspects or features of the cited work (about which the citing author mentions his opinion) and then generate an aggregated aspect level sentiment profile of the paper. I discuss the problem of aspect-based citation sentiment analysis using a recently published manually annotated corpus for complementing the citation aspects with its purpose and polarity for extracting the most frequent citation aspects. I hypothesize this based on the intuitive observation that the occurrence frequency of domain-specific aspects will be strikingly high in a particular domain as compared to other domains. This will facilitate in measuring the aspect-based sentiment strength or intensity of citations (the level of positive and negative citations and aspects). I first extract the aspects from the citation sentences using the linguistics phrase patterns, synonyms and heuristic rules-based approach and then determine the sentiment orientation of the aspect specific sentiment words using the SentiWordNet by considering the words from around the linguistic expression of the aspect. I group different co-referential aspect cue phrases having the same indication or meaning towards the cited aspect based on WordNet synonym dictionary. Afterwards, I employ chi-square based aspect weighting and ranking mechanism to rank aspect and cue phrases. Moreover, I propose a framework for citation subjectivity detection using different machine learning algorithms (SVM, Naïve Bayes, Maximum Entropy, J48 and Random Forest). Different POS n-gram based features ('N-gram after', 'N-gram before' and 'N-gram around') considering bigrams, trigrams and pentagram are employed to analyse the impact of feature selection and extraction methods on citation classification accuracy. The results are evaluated by utilizing standard evaluation measures such as, precision, recall, accuracy and f-measure.

1.4 Research Questions

RQ1. What is the impact of higher order POS and word-based n-grams in detecting the citation sentiment? (RP-1)
To answer this question, I propose a technique for identification of the sentiment of citing author towards cited paper by extracting unigram, bigram, trigram and pentagram adjective and adverb patterns from the citation text. The findings for this research question has been presented in section 4.1 of Chapter 4.

RQ2. How the type of n-gram model effect the aspect-based sentiment classification? (RP-1)
In determining the importance and relevancy of paper for the researcher based on different aspects of the cited work such as technique, corpus, method, task, concept, measure, model and tool etc, I extracts the aspects from the citation sentences using the pattern of opinionated phrases around the aspect and consider the words before, after and around the aspect using n-gram based features: 'N-gram after', 'N-gram before' and 'N-gram around'. The findings for this research question has been presented in section 4.2.2 of Chapter 4.

RQ3. How can different machine learning techniques contribute to aspect-based sentiment detection? (RP-3)
To evaluate the performance of the proposed technique I use dierent ML classiers like SVM, Naïve Bayes, MaxEnt, J48 and Random Forest for comparative domain corpora. The evaluation for these techniques against citation corpora has been presented in section 4.2.2 of Chapter 4.

RQ4. What is the pattern of citation sentiment distribution across rhetorical structure of research paper? (RP-2)

To answer this research question, I study the distribution and pattern of citation contexts around the IMRaD Structure and its impact on the citation sentiment. The findings for this research question has been presented in section 4.3.1 of Chapter 4.

RQ5. How varied sized contextual window effect citation sentiment detection? (RP-2)

For answering this question, I explored the dierent values of POS n-gram parameter (ranging from 2 to 5) to discover the best setup covering adjectives, verbs, adverbs and noun phrases. Section 4.2.2 of Chapter 4 describes the findings and results against this research question.

RQ6. What is the effect of citation context on citation polarity and purpose classification? (RP-3)

The next step is determining how sentiment varies over the citation context and how diversied is the sentiment contained in the contextual sentences. Here, I investigate the impact of different sizes of contextual windows on a generic dataset for identifying the citation sentiment. The results pertaining to this research question has been described in section 4.3.2 of Chapter 4.

Chapter 2

Literature Review

2.1 Introduction

The domain of content-based citation analysis explores different aspects like Citation Motivation Classification [9, 43–46], Citation Function classification [4, 5, 7, 13, 15, 22, 47–50], Citation Context Identification, Content Analysis of Citation Contexts, Citation Recommendation Systems, Citation Summarization. I performed a systematic review of literature by searching online databases using keywords like Citation Sentiment Classification, Higher-order n-grams, Context-based Citation Sentiment Analysis etc. and considering only those studies that are relevant to citation domain. The detailed research process is presented in Fig. 2.1. This chapter presents the state-of-the-art literature for those directions of research that are closely related to our proposed work.

2.2 Citation Sentiment Classification

Existing studies consider different techniques for approaching citation sentiment analysis. Researchers are emphasizing citation content and citation context investigation for evaluating the scientific contributions quantitatively. For characterizing the contributions made by the cited paper towards the citing they have

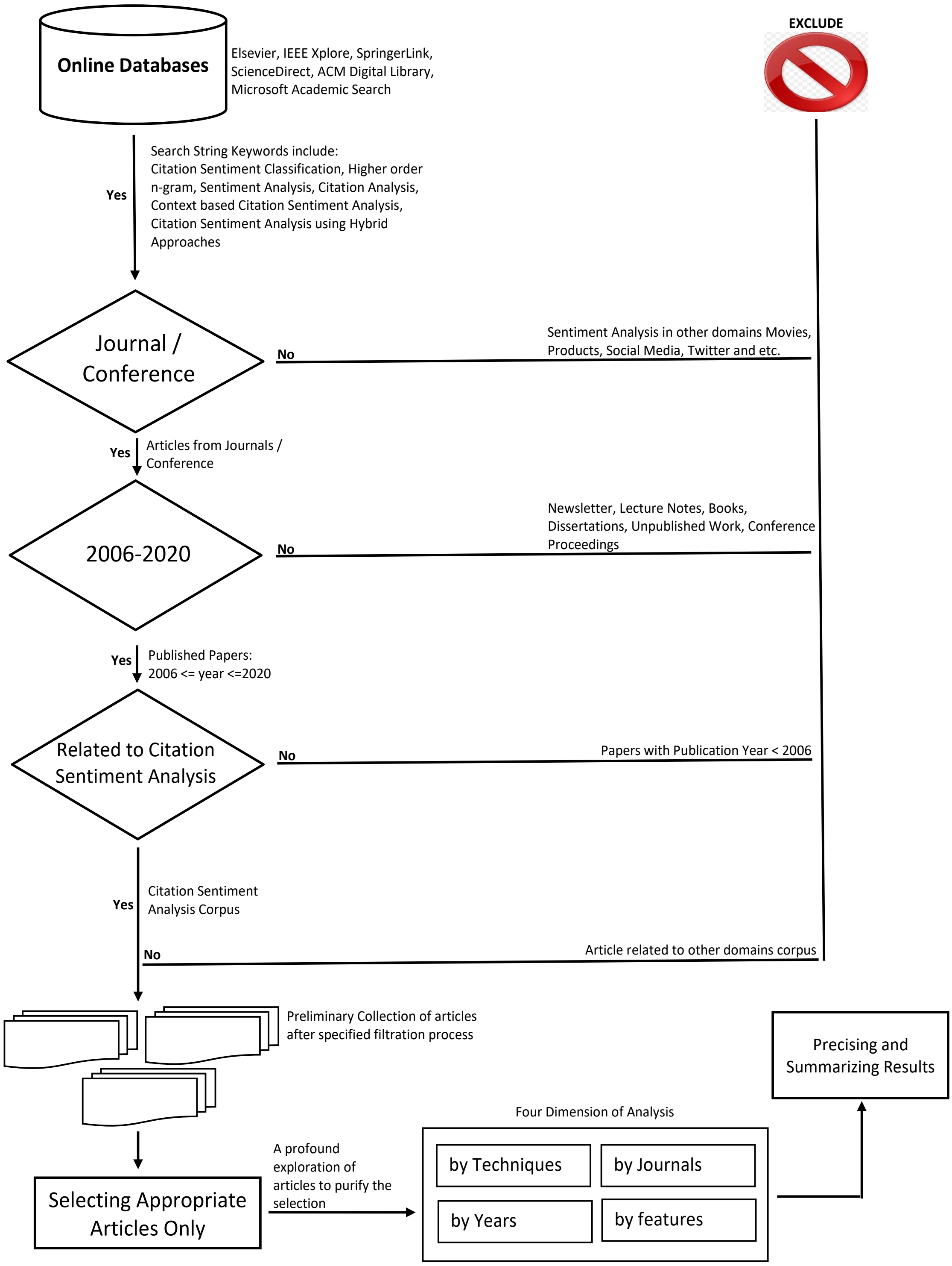

Figure 2.1: Research Process.

exploited different text mining and natural language processing techniques. Here, I summarize the relevant literature for sentiment classification of the citations for the scientific papers. [1] focused on the automatic sentiment polarity identification in the citation text using different word-level linguistic features. The author has used features like n-grams (n=3), dependency relations, negation features, scientific lexicon and sentence splitting in the SVM framework. SVM categorized the citations into three different classes, i.e., positive, negative and neutral. The results further ascertained that trigrams and dependency relations offer robust results in this regard and outperform the scientific lexicon and sentence splitting features. [5] intended to develop an automated method for citation sentiment detection using machine learning techniques and linguistic clues. In their preliminary work, the authors have presented a technique for citation text classification based on support vector machines using n-gram (unigrams and bigrams) word statistics as a feature vector. The proposed citation sentiment classification technique categorizes the text into two categories, i.e., positive and others. In future work, they plan to improve their method by using a denser ground-truth dataset and enriching the feature set by considering more input features. [7] contemplated a technique to generate the structured summary of the research paper based on the citation text in citing papers. They have classified the citation text using multi-label classification into one or more of five different classes i.e., summary, strengths, limitations, related work and applications. As a baseline, they have used Naive Bayes algorithm while considering the combinations of adjectives, verbs, and n-grams in each class. As per the experimental results, considering adjectives, verbs, and bigrams in combination achieved an average precision of 68.54%. [22] presented a technique for citation bias detection using manual citation sentiment analysis of biomedical research publications. The authors have outlined some differences between the approaches used by biomedical researchers and the automated citation sentiment classification methods. According to their findings, researchers paid a lot of concentration to the citation sentiment aspects like strength and validity and less emphasized on the simple polarity based sentiment. [13] described that many of the existing techniques for the citation sentiment classification correspond to the domain-specific areas like medicine, biomedical, computer science, French humanities articles, etc. In their work, they have classified the citations based on Naïve Bayesian classifier by selecting a window of five sentences around the cited place with an accuracy of 80%. For the experimental purposes, they have used generalized lexica considering the citations from the multiple domains and it establishes the demonstrability of their approach in multiple disciplines. [3] focused on combining text mining and lexical analysis techniques for the elucidation of the author's attitude towards the cited work. They encapsulated the extracted opinions in an objective measure for determining the impact of scientific publications. One of the key findings of their research work is that the majority of the citations were neutral in nature and found a considerable agreement in the utilization of the terminology for verbalizing the sentiments towards the cited work both in terms of positive and negative opinions. As a future study, they want to extend the model by considering different lexical elements and their different combinations such as conjunctions, adjectives, verbs, adverbs and verb-adverb combinations. [15] focused on the textual (cue phrases, no. of cue phrases, etc.), physical (location, popularity, density, etc.) and syntactical features in establishing a connection between the citations function and its polarity. They have classified the citation sentence into four different types: background, fundamental idea, technical basis, and comparison by using different supervised learning classifiers including BayesNet, NaiveBayes, SMO, J48 and RandomForest. They have proposed an ensemble-style self-training based classification model and performed their experiments on citations extracted from computational linguistics papers. The analysis of the results shows the efficiency of using the proposed feature set when combined with syntactic features extracted from part-of-speech (POS) tags. [16] used the classic faceted annotation scheme of Moravcsik and Murugesan (1975) and applied the Standard ME (Maximum Entropy) classifier. They classified the citation sentences into four different aspects to decide whether the cited work is (1) an idea or a tool (2) accurate or faulty (confirmatory or negation) (3) fundamental or perfunctory (4) based on or alternative work. The contextual length of citation sentences consisted of 1 to 3 sentences. A number of features like unigrams, sentence location, linguistic features, comparatives and a lexicon consisting of positive and negative sentences are used in a combinatory fashion. From the four annotation labels, one is used for the sentiment polarity labelling positive instances as confirmative and negative sentences as negational. However, there is no neutral class. One of the key findings of research exploration is that the majority of the citations were neutral in nature with considerable agreement in the utilization of the terminology for verbalizing the sentiments towards the cited work both in terms of positive and negative opinions. A future direction for work could be to extend the model by considering different lexical elements and their different combinations such as conjunctions, adjectives, verbs, adverbs, and verb-adverb combinations. As per the findings, there is still plenty of room for further research and development in the domain of citation sentiment analysis. Further, modelling the negations in the citation sentences for the performance improvement could be a possible direction which will help in exploring longer citation contexts. This can also help in removing the dependency on manual annotation and appropriately tagging and recognising the citation categories. With reference to the citation sentiment analysis there are some challenges as well more specifically where the datasets are non-standardized. This could be the result of presence of multiple citation formats and reference styles. The size of the citation context in which the citation anchor exists is also an essential and challenging task to explore. The presence of noise in the training datasets, small dataset and data annotation anomalies also add on to these challenges. To stimulate the citation sentiment research activities based on natural language processing and machine learning based approaches, the contributor authors in this domain should make their coding and annotated citation datasets readily available. The citation sentiment analysis domain can be further augmented with authors affiliation with the institutions, ranking and journal indexing to give a more aggregate level of citation analysis for provision of nuance to research studies in bibliometrics. In regards to better instigate the impact of the research study and its sentiment impact, another future direction is considering the total time elapsed from the publication and sentiments received over the period of time.

TABLE 2.1: A comparison of different context based citation sentiment classification techniques

Author	**Year**	**Categories**	**Features**	**Classifiers**	**Corpus**	**(Accuracy %)**
Athar [1]	2011	Positive Negative Neutral	N-grams(1-3) Dependency Relations Negation Features Scientific Lexicon Sentence Splitting POS Tagging	SVM	8736 Citations from 310 Research Papers	N-grams (1-3 grams): macro-F: 0.597, micro-F: 0.862
Dong & Schäfer [15]	2011	Background Fundamental idea Technical basis Comparison	Textual Feature Physical Features (location, popularity, density, AvgDens) Syntactic Features	Supervised Methods (BayesNet, NaiveBayes, SMO, J48 and RandomForest) Ensemble Learning Model	N = 1768	BN: 0.64 NB: 0.66, SMO: 0.64 T The supervised classification models NB and BN have shown the best performance.
Tandon & Jain [7]	2012	Summary Strengths Limitations Related work Applications	N-grams Verbs Adjectives Verb & Adjective Combinations	Naive Bayes	500 Citation Contexts from 30 Research Papers	Adj.: 65.54, Verb:66.30, Adj.+Verb: 67.48, Adj.+Verb + Bigram: 68.54
Jochim & Schutze [16]	2012	Four different facets (one facet is for sentiment polarity)	Unigrams Sentence Location Linguistic Features Comparatives Lexicon of positive and negative sentences	Maximum Entropy classifier	2008 Citations	Accuracy: 89.7%
Butt et al. [13]	2015	Positive Negative	Sentiment Lexicon	Naïve Bayes	Dataset = 150 Papers	Naïve-Bayes: Precision:0.75
Kim & Thoma [5]	2015	Positive Others	N-grams (unigrams & bigrams)	SVM	2,665 CON Sentences	Unigram: 84%, Unigram + Bigram: 86%

Table 2.2: A comparison of different context based citation sentiment classification techniques

Author	**Year**	**Categories**	**Features**	**Classifiers**	**Corpus**	**(Accuracy %)**
Khalid et al. [49]	2019	Key Citation Topic Modeling Citatin Context	Model Compound-Noun Occurrence Word Preposition Word-Sense-Disambiguation Context Window Size	LDA Algorithm	Cited Articles (113) from ANN Corpus	Adj-rand: (0.425) V-measure: (0.425) Norm-mutual-info:(0.432)
Kilicoglu et al. [51]	2019	Positive Negative Neutral	N-gram Features Sentiment, Lexicon Features, Structure Features, Dependency Features, Rule-based Features	Support Vector Machines (SVM) Convolutional Neural Network (CNN) Bidirectional Long Short-Term Memory (BiLSTM)	285 Discussion Sections (4,182 Citation Papers)	Accuracy: 0.882 Micro-F: 0.721
Yousif et al. [52]	2019	Citation Sentiment (Positive, Negative, Neutral) Citation Purpose (Criticizing, Comparison, Use, Substantiating, Basis, Neutral)	NB with Syntactic Features, SVM with Features, TF-IDF Embedding, CNN with Embedding, Single task (LSTM, BiLSTM, RCNN) Multitask (LSTM, BiLSTM, RCNN)	Multitask Learning Model Using Convolutional & Recurrent Neural Networks	Dataset1: 3568 Citation Contexts Dataset2: 1768 Instances	Precision: 92.10% Recall: 84.87% F-score: 88.34%
Cohan et al. [45]	2019	METHOD RESULTCOMPARISON BACKGROUND EXTENSION FUTURE MOTIVATION USE	Citation worthiness Section title Multitask formulation	Neural Scaffold	ACL-ARC (186 Papers) SciCite (6627 Papers)	Average F1 Score: 84.0

2.3 Context based Citation Sentiment Classification

Up till this point, I have presented the state-of-the-art literature on citation sentiment detection based on only the citation sentences. But sometimes only the citation sentence is not enough in detecting the sentiment orientation of the citation because of the objective nature of the citation sentence. The context-based citation sentiment analysis includes considering both citation content in a window near citation reference and citation position [32, 47, 53–57]. As sometimes only the citation sentence is not enough in detecting the sentiment orientation of the citation because of the objective nature of the citation sentence. The context of the citation also impacts its sentiment orientation [9, 14]. I can categorize the citation context both at the syntactic and semantic levels [58]. The syntactic citation context refers to the citation mention (in how many research papers the article has been referred to) and citation location (at which location i.e., rhetorical structure this citation mention has been done in the citing research paper) [18]. The attributions of cited paper, citation function and sentiment analysis are performed based on the semantic citation context [42]. New researches demonstrate that combining different citation contextual features exhibit a better potential for accurately evaluating the citation contribution than merely the citation count. However, a major task in the context-sensitive citation analysis is the extraction of context which have primarily been applied in citation summarization, author ranking, impact evaluation and sentiment analysis [59–62]. In this section, I present some recent work that demonstrates the importance and difficulty in extracting and analyzing the citation blocks which are spans of citation sentences encompassing one or more contextual sentences. One of the earlier attempts for determining the span of citation text has been made by [63] who applied it for retrieval of cited documents based on the citing statements. A set of handcrafted heuristic rules are applied for the extraction of citation sentences within a window of +/- 2 contextual sentences around by using different cue words. I consider the approach of [63] as a baseline. A varied size of contextual citation sentences has been reported in literature for citing the target paper comprising of a few words to a number of sentences [17–19, 64]. [32, 65] utilized the symmetric window approach of context extraction considering 50 words before and after and 300 words before and after respectively. Similarly, [66] applied a sentence selection approach based on a window of contextual citation sentences consisting of citation sentence, 1 sentence before and 1 after. [67] annotated a corpus considering a 4 sentence window, 2 sentences before and after and proved by numerical evidence that each citation sentence is linked to only a few adjacent sentences. The proposed technique investigated the usefulness of considering such contextual sentences for generating fluent scientific surveys by combining bibliometric and summarization techniques. Furthermore, [68] considered the length of citation context as 5, incorporating 2 sentences prior to the citation mention and 3 forthcoming sentences. Another approach for citation block determination is proposed by [58] by using textual coherence features. The proposed technique demonstrated promising results by considering the citation block starting from the citation sentence along with the forthcoming sentences by using SVM and CRF classifier. In this study, I initially consider the sentence containing the citation tag and further extend the citation block with contextual sentences in a window size of +/- 3 for studying the impact of contextual sentences on citation sentences in the domain of citation sentiment analysis. It has been further argued in the literature that identifying the proper citation context would be a difficult task because of its domain-specific nature without human intervention. The analysis of the state of the art establishes that considering a fixed window size and ignoring the coreference will result in introducing false positives (FP) and false negatives (FN) which will lead to noise. Although many research studies have reported a diversification in citation function, patterns, distribution and motivations, little attention has been given to analyzing the pattern of citation sentiment distribution over paper structure. In order to fill this gap, I consider citation sentences to quantitatively evaluate the behavior of citation sentiment mention over the paper structure. [9] used the contextualized citation sentences for scientific sentiment classification. In this extended study, they have demonstrated that sentiment orientation of the citation sentences is not just emerged from the citation text particularly in the case of negative sentiment. Rather it is the context which determines its sentiment orientation. Citation context both explicit and implicit improves the overall performance of citation sentiment detection. They have used the context window of different sizes to determine its impact on the citation classification. However, the annotated context window consisting of 4 sentences have increased the overall classification performance by increasing the number of negative citations. They have proved that if we ignore the citation context it would result in unidentified citation classification specifically in the case of criticism. [14] developed a method for the generation of citation summary of the research papers based on the context in which the citing paper has cited it. They have used different features like the semantic similarity between both the citing and cited article, contextual sentiment analysis of the citation text and weighted self-citations. Based on the pageRank algorithm which considers the features specified they not only generate the citation summary of the article but also generate the qualitative citation index for the cited article. As a future work, they want to extend the citation quality index considering more quality features. [23] intended to investigate the citation sentiment by applying machine learning techniques on clinical trial paper corpora. The main contribution of this study is the creation of concept-level annotated corpora consisting of clinical trial papers and applied machine learning based classifier by using different features like n-grams, sentiment lexicons and problem specified structure features individually and in a combinatory way. For the sentiment detection from the citation text they have considered the citation context. The novelty of this research is as being the first study addressing the automated citation sentiment classification in the domain of biomedical. As future work, they plan to enhance their approach by employing methods to handle imbalanced data and extracting more informative context features. [4] worked on an annotated corpus and put forward a citation classification scheme considering the function, polarity and impact. They compared the classification results with the survey responses from the authors. The analysis of the results established a strong correlation between the author's feedback and classification results. The authors applied and tested different machine learning algorithms and obtained the best results with SVM trained with SMO. As an extended work, they want a more detailed impact categorization by taking into consideration all the function classes. [17] did the binary classification of the citations based on several features extracted from the full text of the research articles. They classified the citations into two categories either influential or non-influential and discussed that citations are not equal. The main aim was the identification of citation function and detection of the influential effect on the citing article. The authors suggested a framework for the identification of influential citations based upon their repeated use in a paper, the similarity between the citing and cited articles, the citation context in the citing paper and the location or position of the citation in the citing paper. Based on these features they predicted the presence or absence of an influential relationship between the citing and cited paper (the citation edge) using logistic regression and support vector machines. According to the key findings of the paper, the number of times a paper mentions or cites the cited paper was one of the best factors for predicting the influence of the reference. [18] estimated the strength value i.e., the importance of the citation for distinguishing different types of citations. They investigated that simply citation count is not enough in accessing citation relationships and in this regard citation importance is of great use. Therefore, the authors presented a regression method for automatic estimation of the citation strength value and established a good correlation between the estimated strength values and human labelled values. On the basis of a newly created dataset, they hypothesize that all the citations referenced by the paper are not equally important and some of them are more important as compared to others. In future, they want to improve the estimation performance by considering more features from the co-citation and bibliographic coupling. [19] established a faceted citation link classification utilizing textual features, extra-textual features and network structural features. Using support vector machines, they classified the results into three mutually exclusive categories, functional, perfunctory and hard-to-tell (a class for ambiguous cases). They manually annotated the data considering the citation contexts consisting of 3 sentences around the citation text.

Table 2.3: A comparison of different context based citation sentiment classification techniques

Author	Year	Categories	Features	Classifiers	Corpus	(Accuracy %)
Athar & Teufel [9]	2012	Positive Negative	N-grams of length 1 to 3 Dependency triplets	SVM	1,741 Citations	F Macro: 0.731, F Micro: 0.871
Kazi & Patwardhan [14]	2015	Semantic similarity Sentiment analysis Weighted self-citations	pageRank Algorithm	541 Nodes & 659 Edges		
Xu et al.	2015	Positive Negative Neutral	N-grams(1) Sentiment lexicons (2) Problem-specified structure features (3)	SVM	4182 Citations	(1): 0.853, (1) + (2): 0.869, (1) + (3): 0.868,(1) + (2) + (3): 0.870
Hernandez-Alvarez & Gomez [4]	2015	Function(based-on, supply,useful, acknowledge, debate, Hedges) Polarity (Positive, Negative, Neutral) Impact (Negative, Perfunctory, Significant)	Citation Location Citation Mention Count Most citations positive or neutral	SVM	2092 Citations	Ave Precision: 0.929
Zhu et al. [17]	2015	Influential Non-Influential	Count based features Similarity based features Context based features Position based features Miscellaneous features	Logistic Regression Support Vector Machines	3,143 Paper–Reference Pairs	Accuracy: 89.7%
Wan & Liu [18]	2014	Strength value Importance	Occurrence number Located section Time interval Average length of citing sentences Average density of citation occurrences Self-cited or not	Logistic Regression	820 Citations	Mean squared error (MSE): (support vector regression) :0.5419
Xu et al. [19]	2014	Functional (utilize, support or criticize) Perfunctory Hard-to-tell	Textual features Extra-textual features Network Structural features (author relationship, paper relationship & centrality measures)	SVM with linear kernels	1,185 Citation Contexts	Accuracy: 0.84

Table 2.4: A comparison of different context based citation sentiment classification techniques

Author	Year	Categories	Features	Classifiers	Corpus	(Accuracy %)
Vyas et al. [69]	2020	Positive Negative Neutral	Parts-of-Speech Tags Dependency Relationships wvCNN_random wvCNN_non-static	Deep Learning (wvCNN, Oh-CNN, Oh-biLSTMp)	Dataset1:8925 Dataset2:6567 Dataset3:2164 Dataset4:1874 Dataset5:15492 Dataset6:4038 Dataset7:2164 Dataset8:1874 Dataset9:4038	Elsevier (Macro-F1:67.66) ACL (Macro-F1: 77.32) Augmented (Macro-F1: 69)
Mercier et al. [70]	2020	Sentiment(Positive, Negative & Neutral) Intent Classification (Result, Method & Background)	ImpactCite (XLNet-based solution)	CNN LSTM RNN	CSC-Clean corpus (Class 1: 1491, Class 2: 3154 & Class 3: 6375)	Micro F1: 88.13 Macro F1: 88.93
Nazir et al. [71]	2020	Positive Negative Neutral	Content Similarity Citation Count Section-wise In-text Citation Weights	Support Vector Machine Random Forest Kernel Linear Regression	Dataset1:465 Dataset2:311	Precision:0.84
Aljuaid et al. [72]	2021	Positive Negative Neutral Cosine Similarity	Term Frequency-Inverse Document Frequency (TF-IDF)	SVM KLR Random Forest	D1: 465 annotated citation pairs D2: 488 paper citation pairs	F-measure: 0.83

2.4 Aspect based Sentiment Classification

Aspect based or fine-grained sentiment classification is a fundamental problem in sentiment analysis, which has received considerable attention during the past few years. Various algorithms, methods, systems and approaches have been proposed in the literature for aspect extraction and aspect-opinion determination in different domains like movie reviews, product preferences, services, marketing campaigns, blog posts, news, political movements, social events etc. [73–79]. In this section, the critical review of some state-of-the-art techniques for aspect-based sentiment classification is presented as the proposed work focuses on aspect identification and sentiment extraction from citation text. An approach based on association rule mining, SentiWordNet and dependency parsing is proposed for aspect level sentiment analysis of the Twitter dataset [80]. The performance of the technique is evaluated by comparing the results on the dataset of the Hate Crime domain and the benchmark dataset. The proposed technique identified the most frequent explicit aspects implying Association Rule mining, whereas extracts the implicit aspects using dependency parser. SentiWordNet lexicon is utilized to assign the sentiment polarity score for each extracted opinion phrase. The approach achieved an overall accuracy of 81.58% on the Stanford Twitter Sentiment Dataset. Furthermore, they presented an unsupervised method that performed the user level sentiment analysis considering learning-based and augmented lexicon-based techniques on the Twitter dataset. The proposed method identified the patterns of a user's typing habits along with fluctuations in their sentiment expressions using an unsupervised method. The findings of the proposed technique achieved an accuracy of 81.9% outperforming traditional methods. The justification for combining lexicon-based and learning-based methods is that they help to improve the associated low recall with the lexicon-based method along with eliminating the need for manual labelling in the case of a learning-based method. Considering the importance of lexicon-based methods and linguistic expressions within a window of words around the aspect term, following most similar approaches are worth mentioning. Important work has been done by [76] who considered the linguistic expressions around the identified aspects based on the 'n-gram before', 'n-gram around', 'n-gram after' and 'all phrases method'. The local context has been generated based on the bag of n-words approach for n in 2, 3, 4, 5, and 6 to achieve the final result. The experimental evaluation of the approach depicted that the 'n-gram around' method provided better results as compared to the other implied n-gram methods. Recently, [81] gave a novel approach for predicting the financial behaviour of markets. Both aspect and document level polarity classification of financial news has been performed based on the 1000 online financial news. The semantic relations between the extracted concepts and identified polarity have been determined based on an ontology-driven approach. An instance of the technique [82] for aspect level sentiment analysis using a window of words around a given aspect term, SentiWordNet based aspect-polarity assignment and n-phrase rule for aspect classification have been experimented with and validated by [83]. It has been established in the literature that a large number of features deteriorate the performance of ML algorithms because of the presence of noisy and irrelevant features and high dimensionality of feature space [84, 85]. The polarity of the opinion word is highly dependent on the context in which it is used. This paper intends to classify the review datasets belonging to multiple domains by extracting unigrams, bigrams, dependency features and their composites. They have used information gain (IG) and minimum redundancy and maximum relevancy (mRMR) methods for eliminating the redundant and irrelevant features. The experimental evaluation depicted that composite features combining unigram and bi-tagged features performed better as compared to individual features by using the mRMR feature selection method. Our work is related to these approaches because they have used the n-gram methods for sentiment classification. Contrary to our research work, they considered the ontology for aspect identification whereas our technique focused on POS tag patterns to identify the phrases as aspects. Secondly, our proposed scheme is validated in the citation sentiment domain and attempt to improve the aspect-based sentiment classification using frequency-based, opinion-based, rule-based and aspect-ranking methods. In addition, [78] proposed a way to automatically mine the aspect and opinions integrating NLP techniques, SentiWordNet assisted lexicon-based methods and fuzzy sets for estimating the sentiment orientation. In this approach, three different datasets are used to compare the classification results with Naïve Bayes and Maximum Entropy techniques when used in isolation. The findings depicted that the proposed approach demonstrated more precise (84.24%) and accurate (88.02%) classification results as compared to the Naïve Bayes and Maximum Entropy. Another added benefit is the polarity strength identification in the sentences with respect to the base cases. In future, the authors are aimed at extending the approach in terms of developing a real-time interface using SentiWordNet and SenticNet for searching, tagging and polarity identification dynamically. Future directions for aspect-based citation analysis are rigorously implementing the techniques for implicit and explicit aspects determination and to make this process more domain independent as aspects are mostly domain specific. A possible direction in this regard in the implementation of hybrid techniques for aspect identification which further asserts a need for semantic concept-centric analysis of the aspects and their association with sentiments. This could further pathway to the direction of position labelled aspect context extraction and embedding word conclusions along with their semantic relationships. The process of aspect-based citation sentiment analysis can be further improved by applying two different approaches sequentially instead of integrating them which can demonstrate better results as compared to individual approaches. Considering the impact of two step approach along with word embedding would further help in better predictability of aspects and sentiment association. Better methods for aspect-sentiment summarisation like correct pronoun and negation handling and discourse coherence can be utilised for better aspect discovery and aspect sentiment association. Another potential future research direction for aspect-based citation analysis is mapping implicit citation aspects to explicit aspects and author-specific aspect sentiments. For this, already identified opinion chunks from the citation sentences using the ConceptNet and Similarity Index can be leveraged. This will be coupled with performing a weighted citation analysis by assigning weights to each citation mention based on their appearance in the cited research paper.

TABLE 2.5: A comparison of different aspect-based sentiment classification techniques

Author	**Year**	**Categories**	**Features**	**Classifiers**	**Corpus**	**(Accuracy %)**
Zainuddin et al. [86]	2016	Positive Negative	SentiWordNet lexicon	Association Rule mining	Twitter dataset (Hate Crime domain)	Accuracy: 81.58%
Er et al. [77]	2016	Positive Negative	Learning-based Augmented lexicon-based techniques	Unsupervised method	Twitter dataset	Accuracy: 81.9%
Penalver-Martinez et al. [76]	2014	Positive Negative	SentiWordNet sentiment word lexicon Ontology based opinion mining	n-gram before n-gram around n-gram after all phrases method	Movies dataset	Accuracy: 75.82
Salas-Zarate et al. [81]	2017	Aspect & document level polarity classification	Window of words	Ontology driven approach n-gram before n-gram around n-gram after	1000 online financial news	N-gram around: precision: 81.93%, recall: 81.13% F-measure: 81.24%
Appel et al. [78]	2016	Aspect and opinions	NLP techniques SentiWordNet	lexicon based methods, fuzzysets NaïveBayes Maximum Entropy techniques	Twitter dataset Movie Review Dataset	Precision: 84.24%, Accuracy: 88.02%
Al-Smadi et al. [30]	2018	Aspect based sentiment analysis	lexical, word, syntactic, morphological, semantic features	supervised machine learning RNNSVM	Hotel reviews	Naïve-Bayes: Precision:0.75
Wang et al. [87]	2019	Subjective/Objective Sentiment Sentences Positive/Negative Emotional Polarity	Syntactic Structure, Linguistic patterns	Conditional random fields (CRF), Word2vec	ACL Anthology Network 3500 Citation text fragments and 14,000 sentences	Precision:90.13 Recall:86.67 F-Score:93.82

TABLE 2.6: A comparison of different aspect-based sentiment classification techniques

Author	**Year**	**Categories**	**Features**	**Classifiers**	**Corpus**	**(Accuracy %)**
Chakraborty et al. [88]	2020	Appropriateness Clarity Originality Empirical/Theoretical Soundness Meaningful Comparison Substance Impact of Ideas/ Results/Dataset Recommendation Polarity Classification	Syntactic Structure, Linguistic patterns	Multinomial Naïve Bayes (MNB), Random Forest (RF), SVM, BiLSTM, Google's universal sentence encoder embeddings (FFNN-uni), SciBERT embeddings (FFNN-sci)	ICLR Conference Dataset: 8000 Reviews	FFNN-uni F1-Score:0.71
Wang et al. [87]	2020	Important Non-important citations	Syntactic Features Contextual Information and Bibliometric features	Support Vector, Machine, KNN, Random Forest, Pearson correlation coefficient, relief-F entropy weight method	Data1:106,509 Citations Data2:458 Citation Pairs	Precision:0.9 Recall:0.61 F1: 0.73 AUC:0.91
Kastrati et al. [89]	2020	Aspect Extraction Sentiment Classification	tf*idf	Decision Tree, Naïve Bayes, SVM, Boosting	21 thousand manually annotated students' reviews	Precision: 88.72 Recall: 88.61 F1: 88.67

2.5 Summary

[2] reported a literature survey related to the journal citation sentiment analysis. They have presented a framework for citation sentiment analysis, which consists of citation extraction from the paper, pre-processing of the extracted citations, feature extraction and sentiment classification by application of different machine learning techniques like SVM, Naïve Bayes and Decision Trees. According to the survey study different sentiment classification techniques are in use based on the term frequency, n-grams, negations, and lexicon. But still, there is a penalty of room for further work in the domain of data extraction and generic citation corpus creation. The majority of the researcher in the domain of citation sentiment classification have utilized n-grams [1, 5, 7, 9, 16, 19]. Much of them have used n-grams of varied sizes ranging from 1 to 3. [7] achieved the best results with bigram for multi-label classification. [1] achieved better results using trigrams for a two classed citation sentiment classification. Both unigram and bigrams are good in detecting the citation sentiment. Unigram ensures the data coverage and bigram gives the maximum coverage on sentiment patterns. It can be seen from the results that there is no consensus on which feature is the best one. So, the question which arises from this discussion is that which n-grams give the best results? Either it is the unigram, bigram, trigram or some higher order n-gram. Another observation on the state of the art is that almost all the approaches utilized lower-order n-grams maximum up till trigrams for their experimentation. Higher-order n-grams other than trigrams have not been explored in the existing literature to considerable levels for citation classification. Considering the above observation as a research gap, so in this thesis, I want to explore if higher-order n-grams i.e., trigram or pentagram can also contribute to the result. In another survey [35] explores the sentiment, polarity, and performs the function analysis of the citations. As per the findings, there is still plenty of room for further research and development in the domain of citation's sentiment analysis. One of the key challenges for citation sentiment analysis is the non-availability of sufficient size corpus along with standardized annotation schemes. A possible future

direction for work is to improve citation indexes based on the citation influence on the citing article. This would be done by assigning more weights to the influential references and build-upon articles and assigning less weight to criticized ones. A number of studies have highlighted the issue of hedging in citation sentiment analysis [1, 4, 5, 24]. Hedging is actually the avoidance of expressing negative sentiment towards the cited papers by the citing authors. So there is a need to devise effective techniques for hedging detection and for this reason linguistic clues can be used. Similarly, more abundant objective or neutral citations make the citation classification results somewhat skewed towards the neutral class. There is a need for different techniques to address this issue. The analysis of the literature depicts that most of the citation classification techniques are domain-specific and based on supervised machine learning algorithms. Support vector machines, Naïve Bayes and Maximum Entropy are the most frequently used classifiers. [90] summarized the ongoing research in NLP driven citation analysis for scientometrics. They emphasized the importance of the citation context in classifying the implicit citation sentences which don't constitute any explicit anchor to the target paper. For future research in this domain, they have ascertained the need for building integrated datasets, frameworks and pipelines for evaluating different features and tasks together. Another limitation of the existing research is the usage of a fixed-sized contextual window. All the existing studies have been performed on different datasets. Therefore, the results of these studies are not comparable and cannot be generalized because of the variability and diversification of the used features, approaches, measures and corpora. I have not found any study considering different contextual window sizes on a generic dataset for determining the optimal contextual window for citation context identification. So, one possible research direction will be optimal citation context detection, considering contextual windows of different sizes. This research study intends to apply machine-learning techniques for automatically identifying the aspects or features of the cited research work. A significant amount of research work has been done in the domain of citation sentiment classification [4, 7, 16, 28, 38, 40] while relatively smaller amount of work has been done that primarily focuses on recognizing citation aspects and its sentiment expression. Most of the existing studies on citation sentiment analysis establish a relationship between cited and citing work (e.g. background, comparison, acknowledgement, and weakness, positive, neutral) but are not inclined to scrutinize the fine-grained aspects of the cited work. Here, I review some of the latest works related to our research and articulate how our work is contrary to them. [28] presented an approach to automatically summarize scientific papers. A trainable technology is adopted for addressing some of the challenges in summarizing the scientific paper using the citation network. The technique presented in this paper performs the following tasks: (1) For each reference in the Citing Paper (CP), it first identifies the span of the citation text which is referred to as Citance; (2) For the citation text, it identifies the predefined set of facets (Aim, Implication, Hypothesis, Method or Results) of the Reference Paper (RP) it corresponds; (3) Later, it generates a structured summary of the reference paper based on the extracted aspects from the cited text which spans up to 250 words. The proposed technique is the supervised one that combines evidences from several sources. For the facet identification problem, SVM classifier is applied using 10-fold cross-validation. The scheme has focused on a fixed set of facets or aspects that have already been referred to as citation function or purpose in the earlier research studies. [91], proposed a technique for summarization of related work from the scientific papers. The applied scheme first extracted the citation sentences by combining different techniques like regular expression-based approach, evidence-based approach, co-reference system and additional extraction rule. For the classification of citation sentences into the rhetorical categories (problem, method and conclusion) n-grams, thematic words, cue phrases, sentence length and term frequency are used as a feature vector. Naïve Bayes, Complement Naïve Bayes and Decision Tree machine learning classifiers are used for the learning of the classification model. Analysis of the results describes that for the extraction of citation sentences task combinatory approach shows good performance. Cue phrases and thematic words in combination perform better as compared to full features. The work of [92] is similar to the approach in terms of evaluating the presence of linguistic patterns based on the n-grams in the citation context along with their relationship with the rhetorical structure of the article. The objective of the study was the in-depth semantic analysis of the scientific article. The findings of the study depicted the presence of the most frequent linguistic pattern governed by the rhetorical structure of an article. The presence of verbs in the n-grams is utilized to identify the citation functions and similarity matching in full texts. Although, our proposed framework share similarity in terms of (using the linguistic patterns for citation analysis). However, our scheme is different in two ways. First, the previous work is based on the occurrence of verbs in the n-grams and their distribution in the rhetorical structure of the article whereas I focus on the presence of nouns for the identification of the aspects from the citation sentences. Secondly, I perform the aspect-based sentiment analysis by extracting the opinion phrases (adjectives JJ and adverbs RB) from the citation sentences whereas earlier work just identifies the citation sentences that can be annotated with citation function. The analysis of the relevant literature depicts that to determine the correct size of the sliding window i.e., the value of n, for performing the word n-gram based sentiment analysis on the citation dataset is the relevant area of exploration and experimentation. Further, I intend to explore which n-gram method i.e., 'n-gram before', 'n-gram after' and 'n-gram around' is more deterministic with reference to aspect-based citation sentiment classification. Secondly, it is established in literature that citation context is an important source of information for determining the citation sentiment. Considering only a fixed-size window of the citation sentence and not considering the context will result in false negatives and positives. Citation context is considered to be n tokens or n sentences before and after the citation sentence including the citation sentence plus n contextual sentences before, after or around it. Therefore, in this study, I aim to explore the impact of contextual polarity on the citation sentences within a varied size sentence window up and down the citation sentence.

Chapter 3

Research Methodology

3.1 Introduction

In this section, the proposed research methodology is described which is based on the following tasks: (1) extraction of the n-grams from the citation text (2) identification of the aspects which will be analyzed (3) extraction of the opinionated phrases for each aspect in the citation sentence (3) identification of the polarity expressed about a particular aspect in the citation sentence (4) classification of the citation text as positive, negative and neutral (5) evaluation and comparison of the higher-order n-grams based citation classification with the commercial counterparts.

3.2 Extraction of N-grams

The citations text is segmented into sentences and the part-of-speech tagging is performed on the gold standard dataset [1] for annotating each word and symbol. A central and important aspect of sentiment analysis is to select a good feature representation. Each citation sentence is split into n-grams of different sizes (unigram, bigram, trigram and pentagram). Unigrams are the extracted bag-of-words separated by the spaces. For example, in the sentence 'The proposed method outperformed the class-based model', after removing the stop word "the", the words 'proposed', 'method', 'outperformed', 'class', 'based', 'model' are all distinct unigrams. Further, some of the topic independent words i.e., stop words (I, you, an, the, we, you, my, to, his, for, nor, but, yet, or, etc.) are removed from the dataset to make it suitable for analysis. However, all the traditional stop words are not removed because some of them more specifically the negating words are sentiment bearing. For example, the word "not", "like" is usually considered as stop words but considering the sentence "author do not like the performance of the classifier", here these words have an important sentiment discriminative power. Bigrams, trigrams, and pentagrams are the features consisting of two, three and five consecutive words. For example, in the sentence 'the results are not competitive to the state-of-the-art systems', after removing the stop words "the" and "to", 'results are', 'are not', 'not competitive', 'competitive state-of-the-art', 'state-of-the-art systems' are distinct bigram features. The advantage of using these features is that they are capable of containing some contextual information. Considering contextual information in terms of higher-order n-grams will not only help in capturing short distance negations but also facilitate in capturing subtle meanings in the form of implicit negations. Therefore, I am motivated and focused on using higher-order n-grams features for citation sentiment classification. An example of these features is shown in Table 3.1. For the sentiment analysis of the citation corpus, the proposed approach is not only contemplating the adjectives and adverbs POS tagged n-grams independently in itself but also evaluate their continuous and consecutive word sequences in the form of n-grams. It considers the long adjective, adverb, and their combinatorial phrases, for deducing the subjectivity to demonstrate their appropriateness for citation sentiment analysis. The reason for using bigram, trigram, and pentagram adjectives and adverbs is that they are more sentiment-bearing and reflect a qualitative judgement regarding a piece of text. For pentagrams it uses (JJ/RB—*, *—JJ/RB or *–JJ/RB–*), trigrams (JJ/RB–*, *–JJ/RB, *-JJ/RB-*), bigrams (JJ/RB-*, *-JJ/RB) and unigrams consist of just JJ/RB. Here * denotes the presence of any other part-of-speech tag in the sentence. A detail of the POS tag along with

TABLE 3.1: POS Tag List and Meaning.

Part-of-Speech Tags	**Meaning**
JJ, JJR, JJS	Adjective, Adjective Comparative, Adjective Superlative
NN, NNS, NNP, NNPS	Singular Noun, Plural Noun, Proper Noun Plural
CC, PRP, CD	Coordinating Conjunction, Personal Pronoun, Cardinal Digit
RB, RBR, RBS	Adverb, Adverb Comparative, Adverb Superlative
VB, VBD, VBG, VBN,VBP, VBZ	Verb, Verb Past Tense, Verb Present, Verb Past Participle

TABLE 3.2: Example of various features.

N	**Feature Type**	**Features**
1	Unigram	proposed, method, outperformed, class, based, model
2	Bigram	proposed method, method outperformed, outperformed class, class based, based model
3	Trigram	proposed method outperformed, method outperformed class, outperformed class based, class based model
4	Pentagram	proposed method outperformed class based, method outperformed class based model

abbreviated and detailed form is described in Table 3.2.

Considering the citation sentence "model produces excellent alignment results", the POS tagged pattern for the review will be "model_NN produces_VBZ excellent_JJ alignment_NN results_NNS". For this sentence our trigram will be of the form "excellent_alignment_results", two bigrams "produces_excellent", "excellent_alignment" and one unigram "excellent". The same method will be followed to find the n-grams patterns based on the adverb RB tag. So, the feature set that is given as input to the sentiment analysis program for experiments includes (1) adjectives (2) adverb (3) adjectives + adverb combination (4) n-grams. I have focused on using higher-order n-grams features for citation sentiment classification. For higher-order n-grams, the probability of a POS tag sequence is calculated as the product of conditional probabilities of its trigrams and pentagrams. So if the tag sequence is denoted as t_1, t_2, t_3, ... t_n and the corresponding word sequence as w_1, w_2, w_3, ... w_n, the above-stated fact can be explained with Eq (3.1 and 3.2).

$$L_{min} \leq Length\,(pos\_sequence) \leq L_{max} \tag{3.1}$$

$$P\,(t_i \mid w_i) = f\,(t_{i-4}, t_{i-3}, t_{i-2}, t_{i-1})\,/f\,(t_{i-4}, t_{i-3}, t_{i-2}, t_{i-1}) \tag{3.2}$$

This provides the transition between the tags and helps in capturing the context of citation sentences in terms of higher-order n-grams. The probabilities are computed with the following formula in Eq (3.3):

$$P\,(t_i/t_{i-4}, t_{i-3}, t_{i-2}, t_{i-1}) = f\,(t_{i-4}, t_{i-3}, t_{i-2}, t_{i-1})\,/f\,(t_{i-4}, t_{i-3}, t_{i-2}, t_{i-1})) \tag{3.3}$$

For the sentiment analysis of the citation corpus, the proposed technique have not only contemplated the adjectives and adverbs independently but also have evaluated their continuous and consecutive word sequences in the form of n-grams. It has also calculated the degree of resemblance of each word pattern in the citation data d to the citation text as mention Eq (3.4):

$$resp(dt) = \begin{cases} 1, citation\,vector\,contains\,pattern\,as\,it\,is,\;in\,same\,order; \\ \alpha.n/N, n\,words\,from\,N\,words\,of\,pattern\,appear\,in\,text\,in\,correct\,order; \\ 0, if\,no\,words\,of\,pattern\,appear\,in\,citation\,text; \end{cases} \tag{3.4}$$

3.3 Skip-gram

A generalization to handle the sparsity problem associated with the n-grams is to use the skip-grams by skipping some words in a sequence over arbitrary gaps. A skip-gram is of the form k-skip-n-gram, which is usually an n-gram consisted of some adjacent terms and the value of k determines the maximum number of words or terms to be skipped. In the example, "proposed similar statistical tagging model", "proposed similar statistical translation model", "proposed similar statistical alignment model", is an example of 1-skip-5-gram (a pentagram 5-gram with 1 skipped word i.e., 1-skip-pentagram). Considering the citation sentence "The IBM models produce excellent word alignments.”, the resulting 1-skip-2-gram representation: <The IBM>, <The models>, <IBM models >, <IBM produce>, <models produce>, <models excellent>, <produce excellent >, <produce word>, <excellent word>, <excellent alignments>, <word alignments>. More specifically, if more gaps are allowed between the words this will result in a further increase in the number of extracted skip grams.

As skip-gram predicts the surrounding/closer words (context) in the vector space by being provided with the current word, this approach was utilized for getting the adjective gerunds associated with the noun phrases for identifying the context and sentiment associated to the Noun phrase. So, the intent is predicting neighboring words given central word phrase as the main input. For this, it considered skip-grams and trained the model with different contextual window sizes (2, 3, 5) for considering the neighboring words before, after and around the noun phrase on the citation corpus to compute word embeddings. The objective is the maximization of the likelihood of a word based on other words in the same sequence. The skip-grams was utilized for considering the example of words that exist far away from each other but considering together provide substantial contextual information. For example, considering the word ‘outperform’ as the target work the context can be predicted, it will indicate that there is a high possibility that its context is ‘The technique has really [...] other approaches’. However, if the size of the window is enlarged, the quality of the generated vector would enhance but the computation cost will also be increased.

3.4 Document Level Citation Sentiment Classification

In the sentiment classification module, the unsupervised sentiment analysis method was used to determine the sentiment polarity of the citation text based on the SentiWordNet 3.0 dictionary. After targeting the whole term profile of the citation sentence, it extracted only those terms that are labelled as adjectives, adverbs, and

```
Model: Citation Sentiment Classification
Input: Citation Corpus
Method:
        For each Citation Context
          If Contextual Sentences > 1 Then
               Split context into sentences
               Identify transition connectives
          End If
          Perform POS Tagging
          Extract POS Tag Pattern (i.e., Adjective and Adverb)
        End For
        For each Adjective and Adverb
          Perform Stemming
          Formulate the n-grams of different sizes
          Calculate the Sentiment Score of each tokenized string
          Aggregate for overall Sentiment Score of each citation text based on n-grams
        End For
          Make initial seeds for Sentiment Analysis
          Input n-grams for Sentiment Analysis
          Assign Sentiment Polarity to each Citation text based on n-grams
          Assign Sentiment Polarity of Contrastive Contexts based on Sentiment Strength
          Compare the results of different n-grams
          Compare the results with Manually Classified Citations
          Compare the results with Citations Classified by Commercial tools
Output: Best Classified Citations
```

Figure 3.1: Proposed Model for Citation Sentiment Classification

their n-grams. Computational linguists suggest that adjectives and adverbs mostly represent the sentiment in the sentence. For example, if the citation sentence says "SVM outperformed other classifiers and demonstrated excellent results", here the presence of the adjective "excellent" demonstrates that the author is appreciating the cited work. Similarly, the presence of adverbs further modifies the sentiment orientation of the sentence. For example, the citation sentence "ambiguity classes have been shown to be successfully employed" expresses a more positive sentiment towards the cited research. The sentiment score of each tokenized string labelled as adjectives, adverbs and their n-grams is calculated from the SentiWordNet lexical analyzer. The frequency of occurrence of each n-gram is also aggregated for each of the citation texts. All the corresponding scores of adjectives, adverbs and their combinations are aggregated to obtain the overall sentiment score of each citations text. Then, the document-level sentiment classification was performed to classify the entire citation text into 'positive', 'negative' or 'neutral' classes. The proposed algorithm based on the SentiWordNet incorporating different steps for document-level citation sentiment classification is shown in Fig. 3.1.

TABLE 3.3: Sentiment score of bigram features.

ID	BIGRAM	SENT_VAL	SENT_TYPE
4	the same	0	Neutral
4	same cut	0	Neutral
4	a reasonable	0.45	Positive
4	reasonable starting	0.45	Positive
4	subsequent research	0.30	Positive
10	neglected notable	-0.49	Negative

However, some citations sentences do not appoint any sentiment polarity the reason behind is that they do not have enough clear opinion words. A snatch of determining the sentiment score and classifying the citation text is presented in Table 3.2.

SentiWordNet scores are calculated for the positive and negative phrases found in the citation text and used for determining the sentiment orientation by classifying the citation text to a class with the highest score. Table 3.3 presents a comparison of the sentiment scores and sentiment orientation based on different n-grams features. The analysis of the results demonstrates a clear tendency of improved and strengthened sentiment scores with higher-order n-grams as compared to lower-order n-grams. A pattern of change in the sentiment polarity assignment from the objective class to the positive or negative class can also be seen from Table 3.3.
In case, if there is the contradiction between the citation text consisting of more than one sentence where one sentence is cited as positive and the subsequence sentence or sentences are cited as negative or neutral, the paragraph is first segmented into sub-sentences. Given the contextual citation sentences and their syntactic structure, the role of each word in the sentences is determined. The segmentation is performed merely using the punctuation, interrogation marks, commas and periods along with manually collected keywords e.g., 'because', 'and', 'but', 'however' and 'since' etc. For the transition connectives, the primarily focused ones are 'but' and 'however' because of their higher occurrence frequency.
Followed by the extraction of sentiment bearing phrases (adjectives, adverbs etc.). Then calculating and aggregating the sentiment strength of polarity words in each sentence. To incorporate the sentiment reversal scenario, the sentence with the higher sentiment strength is considered as the final polarity of the citation context.

Table 3.4: A comparison of sentiment classification of n-grams.

ID	Score (n=1)	Type (n=1)	Score (n=2)	Type (n=2)	Score (n=3)	Type (n=3)	Score (n=5)	Type (n=5)
1	0.37	P	0.73	P	1.1	P	1.83	P
2	0	O	0	O	0	O	0	O
4	0.44	P	1.18	P	1.63	P	2.51	P
10	0	O	-0.49	N	-0.98	N	-1.96	N
11	0	O	0	O	0.41	P	0.41	P
17	0	O	0	O	-0.49	N	-1.96	N
51	0	O	-0.5	N	1	N	-1.5	N

In case of contrastive transition and negatives e.g., if a positively cited sentence is followed by the negatively cited sentence, the overall orientation of the citation sentence will be considered as negative because here the citing author intends to hide the sentiment orientation. Because of the fact negation is the most popular structure that results in polarity shifting. Considering the following scenario, “This is classical technique. However, results produced are inefficient”. Nevertheless, the main objective here is building an approach to incorporate the sentiment reversing scenarios. ‘Without’, ‘not’, ‘lack’, ‘never’ and ‘so on’ are some of the corresponding trigger words that are considered in such scenarios. As an initial step, the technique recognized most transitive sentiment reversed citation sentences and decreased the overall sentiment drift to the whole citation context.

3.5 Sentiment Classification using Commercial Tools

The proposed methodology hasn’t only classified the citation text based on the adjective and adverb n-grams, but also find the polarity of the citation sentences using the commercial tools. For this sake, it used two commercial tools, i.e., SEMANTRIA (which is a Microsoft Excel add-in) and THEYSAY (which is an online sentiment analysis tool www.theysay.io). The main objective of using these commercial tools was the validation of the results obtained by the proposed model and to determine the reliability of sentiment analysis tools. There are a lot of commercially available sentiment analysis tools, but I utilized SEMANTRIA and THEYSAY because they are latest, easy to use, easily configurable and applicable to any kind of documents, database, sentences, and phrases. SEMANTRIA performs the multilevel analysis of the sentences incorporating part of speech, assignment of a sentiment score from dictionaries, application of intensifiers and determine the final sentiment score based on the machine learning techniques. Each citation text is assigned with a numerical sentiment score ranging from -2.0 to +2.0 along with a polarity of positive, negative or neutral. Whereas, THEYSAY assigns the sentiment score to the positive, negative and neutral classes in percentages. It offers an in-depth analysis of the text considering different features like POS recognition, humour detection, gender detection, comparison detection, language detection, risk detection, text summarization etc.

3.6 Aspect Detection

The next step in the aspect based citation sentiment analysis is to identify the aspects for analysis. The basic purpose is the extraction of different aspects from citation sentences that are relevant to the cited work. Aspects are the distinguishable facets, attributes or features of the reviewed research paper. For example, corpus density, efficient approach, algorithm, technique, redundant method and so on are different aspects of a research paper. Multiple citation sentences form the overall attitude of the researcher to the cited research but often contains divergent sentiment. For the aspect detection, the technique focused on the NN, NNS, NNP, JJ, DT, VBG POS tags as these represent the nouns, plural nouns, proper noun singular, adjective, determiners and verb gerunds. A single citation sentence can contain multiple aspects of the cited research paper. For example, considering the citation sentence, “The approach used can not be generalized, but the corpus is comprehensive”. In this single sentence, the citing author talks about two aspects of the cited research paper “approach” and “corpus”. In such a scenario, the sentiment rich phrases i.e., verbs, adjectives, adverbs etc. that are related to the noun phrases are extracted using the dependency relations between them. The key instinct here is the noun phrases that more frequently co-occur with opinionated phrases have a greater tendency to be genuine aspects. Each aspect from a citation text appears in its specific context and possesses its sentiment and semantics. An aspect is not affected by all the contextual phrases. For each identified citation aspect, there is a completely different list of saliency words. For example, in the citation sentence, for aspect "approach", the contextual word "comprehensive" is not as important as "generalized". Since sentiment polarity labels are required for each identified aspect, I need to do the segmentation for citation sentences to deal with each of the aspects and in return consider these segments separately for polarity detection. In the above example, the citation sentence can be broken down into {Citation Sentence, Aspect} pair: CiteSent, {Asp1, Asp2} :- {CiteSent, Asp1}, {CiteSent, Asp2}. For each of the citation aspect pairs, a word that best describes the aspect is considered. Then a window of n words around the aspect term is considered which is the segment of attention for that aspect sentiment detection. A citation sentence can contain implicit or explicit sentences. An aspect is considered explicit if it appears in the citation sentence, whereas if it does not appear but its meanings can be implied then it is regarded as implicit. In the above citation sample, "approach" and "corpus" are explicit aspects whereas, determining the implicit meaning or aspect from a sentence e.g., "using NLP tagger is not a piece of cake" is very difficult. In this thesis I am focusing on the explicit aspects of the cited research work.

I employ the Go tagger linguistic parser to perform the POS tagging and to analyse the citation sentences. Similar to previous research studies that consider noun and noun phrases as aspects or features, our technique also extracts noun phrases from the citation sentences that could become the candidate aspects. Each citation sentence is parsed, and the candidate aspect term is utilized to represent it based on the keywords or cue phrases that are specified to describe it. At the next step, it identified the POS tag of the current phrase along with consecutive POS tags before and after the current phrase. It detected different sequential POS tag patterns of varied lengths consisting of bigram, trigram and up-till pentagram. The frequent sequential patterns are used to identify the phrases as aspects. The

Table 3.5: Extracted Phrase Patterns.

Pattern	First term	Second term	Third term	Fourth term
Pattern 1	JJ	NN/NNS/NNP	–	–
Pattern 2	JJ	NN/NNS/NNP	NN/NNS/NNP	–
Pattern 3	RB/RBR/RBS	JJ	–	
Pattern 4	RB/RBR/RBS	JJ/RB/RBR/RBS	NN/NNS/NNP	–
Pattern 5	RB/RBR/RBS	VBN/VBD	–	
Pattern 6	RB/RBR/RBS	RB/RBR/RBS	JJ	
Pattern 7	VBN/VBD	NN/NNS/NNP	–	
Pattern 8	VBN/VBD	RB/RBR/RBS	–	
Pattern 9	VBZ/VBN/VBD	JJ	NN/NNS/NNP	
Pattern 10	JJ	JJ	NN/NNS/NNP	
Pattern 11	RB/RBR/RBS	RB/RBR/RBS	JJ	NN/NNS/NNP
Pattern 12	VBN/VBD	RB/RBR/RBS	RB/RBR/RBS	

candidate aspect patterns can be categorized into three different categories (1) Noun-Adjective: it corresponds to the combination of noun and adjective gerunds in which an adjective describes a noun e.g., the citation chunk "the model has been improved"; (2) Noun-verb: it corresponds to the noun and verb gerund combinations in which a verb describes a noun e.g., the citation sentence "we refined the corpus"; (3) Verb-Adverb: the combination of the verb and adverb gerunds means a verb is being described by an adverb e.g. "technique is used widely". A detailed understanding of the aspect patterns is presented in Table 3.4: The aspect term is qualified based on the following conditions: (1) it must be a noun phrase. (2) the occurrence frequency of the resulting aspect pattern containing the candidate aspect term must be markedly high in the citation corpora; if the extracted noun does not exist in the frequent POS tag pattern, it is not considered as an aspect. (3) the extracted noun phrases are compared with synonyms dictionary created from (Hernández-Álvarez et al., 2016); the nouns not present in the dictionary are ignored; (4) the participation level of candidate aspect term is high based on the chi-square test (2). The main pivot behind this approach is that the noun phrases having a lot of opinions by the reviewers are most likely to be considered as important and distinguishing aspects as compared to those that do not contain frequent opinions. Since an aspect of cited work can be expressed using different words like Corpus, Corpora, Dataset or Data by the citing author I created an aspect vector for different aspects of a cited research work. The intention is to group

Table 3.6: Contingency Table.

	# of citations containing unigram term "w"	**# of citations not containing unigram term "w"**	**Row Total**
# of Positive Citations	f_{11}	f_{12}	$f_{11} + f_{12}$
# of Negative Citations	f_{21}	f_{22}	$f_{21} + f_{22}$
Column Total	$f_{11} + f_{21}$	$f_{12} + f_{22}$	

different co-referential aspect cue phrases having the same indication or meaning towards the cited aspect based on WordNet synonym dictionary. Then assign the one candidate aspect that has a high frequency as the leader of the aspect category. Another example of this particular scenario can be the use of words like technique, method, approach, process or procedure while referring to the methodology aspect group of the cited work. The synonym relationship between different aspects is determined by using WordNet for all the aspect cues. If the relationship between different aspect cues exists, these are grouped in an aspect category and the most frequent one from them is assigned as the category label. After extracting the aspects from the citation corpus with the lexicon-based method, Pearson's chi-square test is applied for ranking different aspects of research work according to the occurrence frequencies. The Chi-square method has been widely used in machine learning literature for feature selection and dimensionality reduction. The same was used in our scenario to find the dependency of an aspect term "w" with respect to each individual class i.e., positive and negative citations. The basic theme is that if the occurrence of a term is high in both positive/negative citation sentences, then it is an indicator to be considered as an aspect term. This has not only ranked the significant aspects by measuring the lack of independence between a unigram term "w" and the classes i.e., number of positive and negative citations but also assisted in the identification of the potential aspect words from the subjective citations. The full description of the algorithm for aspect detection is presented in Fig. 3.2.

At the first step in chi-square test, a hypothesis is formed that whether the candidate aspect term "w" is independent of the positive or negative citation sentences based on its occurrence in both the classes. Then, the contingency table (Table 3.4) is computed for each of the unique unigram noun phrase "w" in the citation corpus where $f_{i,\ j}$ represents the observed or indicated frequency of the corresponding cell in the positive/negative citation set and $E_{i,\ j}$ corresponds to the expected frequency. The chi-square (2) and the expected frequency ($E_{i,\ j}$) are computed by using the following formula as mentioned in Eq (3.5 and 3.6):

$$x^2\,(w) = \sum i, 2 \sum j = 1, 2(f_{i,j} - E_{i,j})^2 / E_{i,j} \tag{3.5}$$

$$E_{i,j} = (f_{i,1} + f_{i,2}) * (f_{1,j} - f_{2,j}) \,/ f_{11} + f_{21} + f_{12} + f_{22} \tag{3.6}$$

The chi-square was calculated for the highest frequency of aspects and cue phrases used in the subjective citations to correlate the number of occurrences of a particular aspect term that the authors comment as positive or negative. The higher the value of Pearson's chi-square, the unigram aspect term "w" is more dependent with respect to both positive and negative citations set. An aspect term is selected as an aspect indicator, if its chi-square value is not less than "2.706", at the significance level of 0.100.

3.7 Aspect Level Opinion Identification

The next step in the aspect based citation sentiment analysis is the aspect polarity identification for each citation sentence. In this regard, the adjectives are extracted both in comparative and superlative degrees (JJ and JJS POS tags) along with adverbs as these correspond to the opinionated phrase. To determine the aspect polarity, we first need to identify the position of the linguistic expression corresponding to the aspect i.e., (NN NNS) and the words around them as the contextual information around the aspect term plays a vital role in determining its sentiment. For each extracted aspect term, a window of n-grams of different sizes constituting the preceding (words to the left) and following (words to the right) few tokens is formulated to extract the context around the individual aspect. This

Model: Aspect detection algorithm

Input: A collection of citation sentences $\{c_1, c_2, c_3, ... c_n\}$;
Set of candidate aspect words & cue words $\{w_1, w_2, w_3, ... w_k\}$;

Method:

Step 1: For each citation sentence, Perform POS tagging;
Extract the noun phrases (i.e., NN & NNS);
Extract POS tag patterns as candidate for Aspects;
IF noun being described by frequent aspect pattern THEN
aspect ← (NN & NNS);
ELSEIF noun in the synonym dictionary THEN
aspect ← noun;
ELSEIF participation level of noun based on the chi-square test (χ^2) THEN
aspect ← noun;
Assign the potential aspect keyword for each citation sentence C;
Update the occurrence frequency for the candidate aspect w, as Count(w);

Step 2: Assign an aspect label to the citation sentence having max count(w),
If there is a tie between aspects, assign multiple aspect labels;

Step 3: For all the identified aspect_cues
WordNet_SynSet = WNSynSet (aspect_cuei)
IF WordNet_SynSet = TRUE then
aspect_category ← aspect_cuei

Step 4: For each candidate aspect, calculate the Chi-square (χ^2) value;

Step 5: Rank the candidate aspect on the basis of χ^2 value;
Add aspects with Maximum chi score on the basis of the threshold value to the seed aspects

Output: Citation sentences with aspect assignment, final aspects

Figure 3.2: Proposed Model for Citation Sentiment Classification

process was performed searching 5-grams forward i.e., 'N-gram after', backward i.e., 'N-gram before' and 'N-gram around' the noun phrase for the occurrence of an adjective, adverb and their combinations. Instead of merely considering the term frequency of the keywords, our proposed algorithm identifies the opinion phrases (adjectives JJ and adverbs RB) from the citation sentences on the basis of identified aspects and later on associates the opinionated phrases to nouns. This is based on the assumption that the aspect-opinion pair containing the aspect phrase and opinion phrase tend to be recurring and found close to each other. It further checked the opinionated phrases in the next two, three or five phrases after the aspect term whether it is in the opinion lexicon or not. If phrases are matched, then calculate the sentiment score and these two words 'aspect' + 'opinion polarity' are put together as an aspect–opinion pair and assign the sentiment polarity to the candidate aspect. If the next two, three or five phrases after the candidate aspect does not contain any opinionated phrases or they are not in the opinion lexicon, and then check the two, three or five phrases prior/before the candidate aspect. If a match, then the sentiment score is calculated, and these two words are combined as an aspect–opinion pair. Similarly, for the all-around n-gram method examine the two, three or five phrases around the candidate aspect. If a match, then calculates the sentiment score and these two words are put together as an aspect–opinion pair. After this, the sentiment polarity is determined using SentiWordNet, which assigns valence scores to opinionated phrases. SentiWordNet associates three numeric scores with each word ranging from 0 to 1 showing its neutrality, positivity and negativity. All these scores are utilised to calculate the final sentiment score. Hence, the positive, negative and objective score of adjective and adverb terms is computed by using the following formula and these belong to the interval [0, 1] deeming to convey a sentiment using Eq (3.7, 3.8 and 3.9) which is them used in the subsequent equations:

$$0 \leq SwnPosScore, SwnNegScore, SwnNewScore \leq 1 \tag{3.7}$$

$$0 \leq (SwnPosScore + SwnNegScore) \leq 1 \tag{3.8}$$

$$SwnNeuScore = 1 - (SwnPosScore + SwnNegScore) \tag{3.9}$$

Each word in the SentiWordNet may have different score based on the sense in which it is used that corresponds to its synonym set i.e., synset and takes the form as shown in Eq. 10. Each synset has a particular sense number indicating its meaning and ranking on the basis of its usage in that particular sense. The sentiment polarity score i.e., positive (SwnPosScore), negative (SwnNegScore) and neutral (SwnNeuScore) for each word is calculated by taking the average score of the sense to which the word was labelled while performing POS tagging. Each word has different senses associated to it. It considered the relevant senses for the adjectives adverb and compute the SynetScore for each term with the help of following Eq (3.10 and 3.11):

$$SynsetScore(w) = \sum \frac{(SwnPosScore(w, s) - SwnNegScore(w, s))}{s} \tag{3.10}$$

$$w_i =< POS, ID, SwnPosSCR, SwnNegSCR, SYNSETTERMS, GLOSS > \tag{3.11}$$

An aspect is considered as positive if SwnPosScore is greater than both SwnNegScore and SwnNeuScore by summing up all the scores of the opinionated words associated with that aspect. Contrarily, if the SwnNegScore is greater than both SwnPosScore and SwnNeuScore, then the sentiment polarity of the aspect is regarded as negative. Finally, it is considered neutral if the SwnNeuScore is greater than SwnPosScore and SwnNegScore. Eq (3.12) is applied to distinguish the polarity of the mentioned aspect. In return, the sentiment polarity and overall sentiment score are assigned to each citation based on the average SentiScore of each opinionated phrase in the citation text, which is computed, with the help of Eq (3.13).

$$class\_label = \begin{cases} if\ max(SwnPosScore, SwnNegScore, SwnNeuScore) = POS \\ if\ max(SwnPosScore, SwnNegScore, SwnNeuScore) = NEG \\ \qquad\qquad\qquad\qquad else \end{cases} \tag{3.12}$$

$$SentiScore(w) = \frac{\sum_{for\, each\, term\, w} SynetScore(w)}{N} \tag{3.13}$$

Where N corresponds to the total number of opinionated phrases i.e., adjectives and adverbs in the citation sentence. The full description of the algorithm for aspect level opinion identification is presented in Fig. 3.3. For example, in Table 1.1, in the first sentence, the opinion terms “similar” and “good” are near to the aspect terms “method” and “results”. Similarly, in the second sentence, the aspect terms “alignment” and “performance” are close to the opinion terms “excellent” and “issue” respectively. The word “similar” in the first sentence corresponds to the category “adjective” as it was labelled as ‘JJ’. There are six different senses for the word ‘similar’ in the SentiWordNet as shown in Table 3.5, four of them correspond to the adjective category and two of them are stated as being nouns ascertaining a differing score each time accordingly. In this particular scenario, four senses for the word ‘similar’ for being used as an adjective are considered. Therefore, the average positive, negative and neutral polarity scores are calculated for the word ‘similar’ that are: SwnPosScore = 0.156, SwnNegScore = 0.156, Model: Aspect level opinion identification algorithm

Input: Citation sentences $\{c_1, c_2, c_3, \ldots c_n\}$; Set of candidate aspect words & cue words $\{w_1, w_2, w_3, \ldots w_k\}$; n-grams

Method:

Step 1: For each citation sentence, formulate the n-grams for each identified noun (i.e., NN & NNS)

Step 2: Extract the closest adjectives (JJ & JJR), verb (VBG) and adverbs (RB)

Step 3: Work on the segment of c_f containing aspect word w_i;

Let the set of opinion words in sf be $\{s_1, s_2, s_3, \ldots s_k\}$;

Count the # of positive and negative appearances for each Term#POS pair (1, -1, 0);

Find the valence i.e., sentiment score for each extracted adjective, verb & adverb using SentiWordNet

Step 4: Assign aspect polarity based on the affirmation of the adjective (i.e., Positive, Negative or neutral)

Step 5: Calculate the average SentiScore of each opinionated phrase in the citation text

Step 6: Assign the sentiment polarity to the whole citation

Output: Citation sentences with sentiment score, feature opinion pairs

FIGURE 3.3: Proposed Model for Citation Sentiment Classification

SwnNeuScore = 0.688 respectively, computed as:

$$SwnPosScore(w) = (0.375 + 0.25 + 0)/4 = 0.156 \tag{3.14}$$

$$SwnNegScore(w) = (0.125 + 0.25 + 0)/4 = 0.156 \tag{3.15}$$

$$SwnNeuScore(w) = (1 - 0.156 - 0.156) = 0.688 \tag{3.16}$$

3.8 Aspect Sentiment Association

For associating, the identified aspects to the sentiment in a citation sentence a set of certain rules are created. Before applying the rules on the citation corpus, the long citation sentences are divided into multiple sentences and only those segments of citation sentences are considered which at least contain an adjective and noun phrase. This resulted in reducing the underlying dataset by around 2.2% because from 8736 citation sentences 192 sentences didn't have either an adjective, adverb

Table 3.7: A Sample Snippet of SentiWordNet Entries.

POS	ID	P SCR	N SCR	SYNSET	GLOSS
a	2381495	0	0.125	similar#4	(of words) expressing closely related meanings
a	2071420	0.375	0.25	similar#1	marked by correspondence or resemblance; "similar food at similar prices"; "problems similar to mine"; "they wore similar coats"
a	1409581	0	0.25	similar#3 like#1	resembling or similar; having the same or some of the same characteristics; often used in combination; "suits of like design"; "a limited circle of like minds"; "members of the cat family have like dispositions"; "as like as two peas in a pod"; "dog like devotion" ; "a dream like quality"
a	1410606	0.25	0	similar#2 like#3 alike#1	having the same or similar characteristics; "all politicians are alike"; "they looked utterly alike"; "friends are generally alike in background and taste"
n	4743605	0.375	0.375	similarity#1	the quality of being similar
n	6251033	0	0	similarity#2 law_of_similarity#1	a Gestalt principle of organization holding that (other things being equal) parts of a stimulus field that are similar to each other tend to be perceived as belonging together as a unit

or noun phrase. The reduction of the dataset helped in primarily focusing on those cases which are more deterministic. In this section, some of the rules are presented with relevant examples:

1. IF there is only one noun phrase (NN) i.e., aspect term and one adjective (JJ) THEN the sentiment will be considered referring to the aspect.

2. IF there are multiple nouns in a citation chunk and only one adjective is used to describe them THEN the adjective will be considered associated with all the nouns. For example, in the sentence "the simplicity and efficiency of parser make it attractive", the adjective "attractive" will be considered a mention for all the aspects.

3. IF there are two noun phrases having the form of noun subject and dependency relationship i.e., <NN, NN, VBP, JJ>THEN the sentiment of the citation sentence will be considered describing the aspect in the noun subject phrase. Example: [machine translation$^{<\text{NN, NN}>}$] have become [predominant $^{<\text{JJ}>}$]

4. IF a citation sentence contains two aspects in the form of <NN, NN, VBZ, RB, JJ>, where the first aspect being the nominal subject and the second from adverbial modifier THEN the sentiment of the sentence will be considered describing the aspect in the noun subject. Example: [two-stage approach$^{<\text{NN, NN }>}$] is [often-beneficial$^{<\text{RB, JJ }>}$]

5. IF a sentence is complex or compound, THEN break the sentence into smaller sentences and analyze the clauses and phrases independently. For example, this sentence "This technique creates a richer indexed source, however paraphrase acquisition is not efficient" will be broken down because the sentence before and after contain both adjective and noun phrases. Then the sentences will be analyzed on the basis of Rule No. 1.

6. IF a sentence contains negation term or phrase THEN the polarity of the sentence i.e., adjective (JJ) used will be reversed. For example, in the sentence "paraphrase acquisition is not new" the use of the negation word "not" will reverse the polarity of adjective term.

7. IF a sentence contains negation term or phrase THEN the polarity of the sentence i.e., adjective (JJ) used will be reversed. For example, in the sentence "paraphrase acquisition is not new" the use of the negation word "not" will reverse the polarity of adjective term.

8. IF a noun phrase marked as aspect is in a noun-noun like <NN, NN>compound relationship with another noun, THEN instead of just considering the single aspect term a composite of both of them will be considered as an aspect. For example, if in the sentence "cluster tree is an ongoing difficult problem", "cluster tree", "cluster" or "tree" is marked as aspect, then in this scenario

the whole expression will be an aspect and subjectivity will be considered associated with it.

3.9 Citation Context Identification

The goal of citation context identification is to identify the contextual sentences that appear within the window around the citation sentence that highlight the important contribution of the cited work. A citation sentence also called the explicit citation sentence is the one that carries an explicit reference to the cited research article whereas the implicit or indirect citation sentences are the sentences that surround the citation sentence. It attached the citation contexts to the citation sentences within a fixed window of three 3 sentences after the citation sentences and 3 sentences before the citation sentence. The underlying assumption for this selection was that the framing sentences which surround the explicit citation may have pertinent sentiment orientation to citation sentence and the sentences which occur further away are less likely to be relevant. Ignoring opinions in the implicit/framing citation sentences and considering only the sentiments associated with the explicit citation sentences would result in losing important information about the cited research paper. However, to decide about the context of citation is a difficult and challenging task as the citation context can be semantically ambiguous and it varies from citation to citation. For this sake, it utilized different term features i.e., different kinds of cue words and phrases whose presence in the sentence becomes a contributing factor for extended citation context. I included this information for considering the neighboring sentences (citation/non-citation sentences) as they may continue to discuss some aspect of cited work and establish a connection with the citation sentence. In this regard, it considered, 1) connecting terms which establish a relationship between different sentences like the author, they, this, these, previous work, nevertheless, furthermore, nonetheless, additionally etc.; 2) I also reckoned the presence of certain shortcoming terms or aspects of the cited work in the forthcoming sentences like performance, dispute, disagree, defect, drawback, limitation, weakness, functioning, accuracy, failure rate, etc.; 3) methodological terms (technique, method, approach, process procedure, framework, we consider, we utilize, we select etc.) in the neighboring sentences epitomized the methodology followed or adopted by the cited research paper; 4) finding or result terms (show, present, propose, highlight, conclusion, outcome, result, consequence, resultant, answer etc.) give a description of the achieved results either by the cited research paper or the citing paper; 5) comparison terms mention the comparative analysis made between the paper and it is extracted on the basis of terms like compared to, correspondingly, regard to/with respect, regarding, in contrast, on the other hand, on the contrary, comparison, improve, outperform, etc.; 6) the occurrence of summary terms like to summarize, summing up, therefore, then, in conclusion, hence, therefore, etc. help in summarizing the cited work.

3.10 Identification of citation section as Citation Context

The distribution of citation sentiment is examined according to the citation location and study how they are related to the citation sentiment. Generally, the citation locations can be classified into four types namely, "Introduction", "Methods", "Results", and "Discussion" usually called the IMRaD Structure. But for this particular study, it has also taken into consideration the "Literature Review" section as a majority of the objective citations pertains to this section and want to study the behavior of objective citations over the contextual sentences. The underlying hypothesis is that if we become aware of the location where the citation sentiment is located, the status and strength of the sentiment mention can be figured out and weight can be assigned to the sentiment due to the varying role of the article section. Each mention of the citation context is annotated to a certain article section by considering the potential variations in the section headings and classified

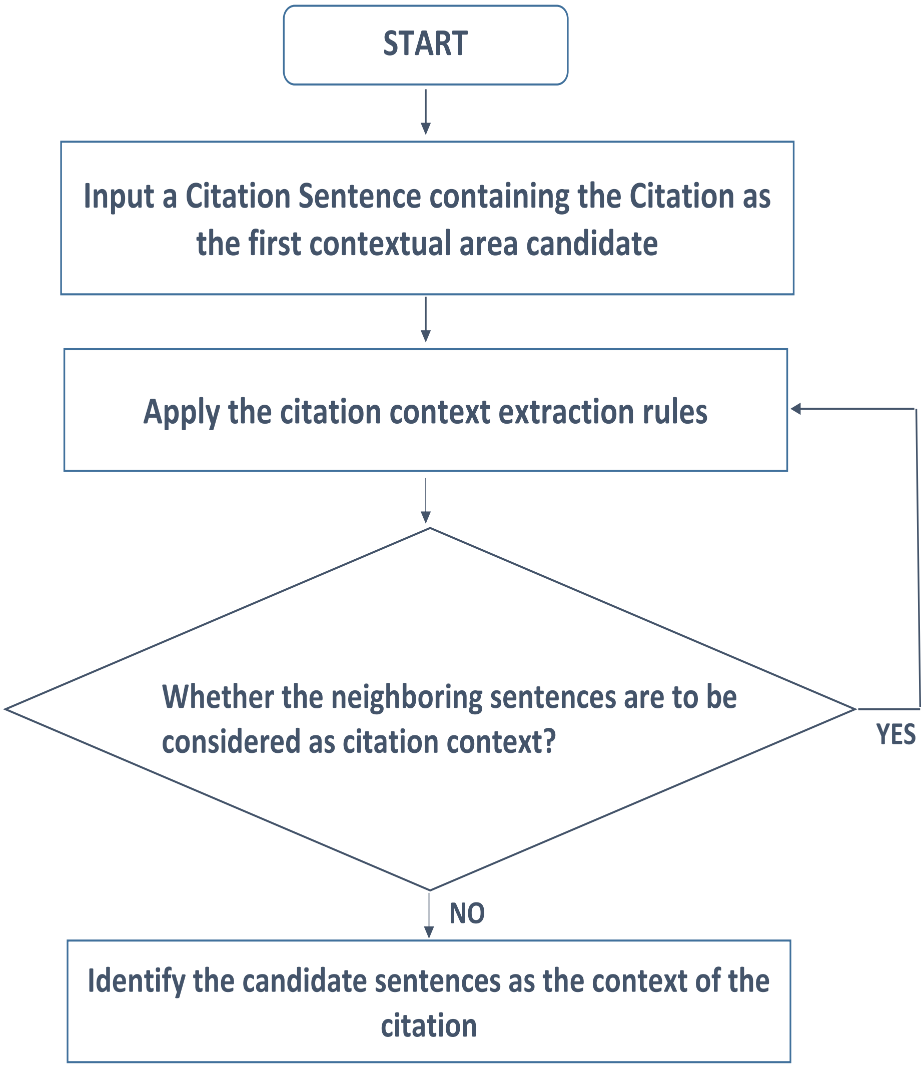

FIGURE 3.4: Flow chart of citation context extraction

them into four main section types. For example, the section "Method" can be represented by different headings such as "Methods", "Method", "Study", "Methodology", "Method and Model", "approach" etc., same as the "Literature Review" section can be expressed by several different titles like "Background", "Literature", "Previous Work", "Related Work", "Framework", "State-of-the-art", etc. A small number of citation contexts in the dataset belong to domain-specific section headings and do not consider IMRaD structure. They were not considered for this particular study. Once the article sections have been identified, I then study the positional distribution of citation sentiment along with the IMRaD structure. It further considers the distribution of citation contexts and the position of these sections along with the IMRaD structure.

3.11 Experimental Setup

3.11.1 Introduction

The experiments are performed on two different datasets belonging to the Computer Science and Bioinformatics domain. The first corpus is a gold-standard dataset, consisted of 8736 citation sentences extracted from 310 research papers from the ACL Anthology Network [1]. This is a manually annotated dataset from which 829 citations are positive, 280 are negative and 7627 are neutral ones. The second dataset is derived from clinical trial papers consisting of the opinionated citation sentences extracted from 285 randomly selected papers [23]. Multiple reference citations are considered as one citation because these represent the same sentiment polarity. The dataset is consisted of 4182 citation sentences extracted from the discussion section of the clinical trial papers arguing the reproducibility of biomedical research. It contains a majority of neutral samples i.e., 3172, along with 702 positive and 308 negative citation sentences. Table 3.8 presents descriptive statistics for both the corpora. This shows the distribution of the classes in the corpora which is highly skewed, with 87% and 76% being neutral citations and only around 14% and 24% are subjective for Dataset 1 and 2 respectively. Both the datasets have been utilized in the previous studies in the domain of citation sentiment analysis that ascertained that most citations are neutral in nature. This is a limitation and well-established fact in the domain of citation analysis as a majority of the citation are cited as neutral. In this scenario, the instances form the majority class (neutral citation sentences) has been thrown away i.e., (under-sample) for having an equal number of labelled samples for each class before training the classifiers. This approach helped in increasing the classifier's sensitivity for the minority class.

TABLE 3.8: Descriptive Statistics of the Corpora.

Sr. No	No. of Selected Papers	No. of Citation Sentences	No. of Positive Citations	No. of Negative Citations	No. of Neutral Citations
Dataset 1	310	8736	829 (9.5%)	280 (3.2%)	7627 (87.3%)
Dataset 2	285	4182	702 (17%)	308 (7.36%)	3172 (75.9%)

3.11.2 Data Collection and Methods

I employ the sentiment, aspect, IMRaD, and extended context to the corpus of Athar et al. which is comprised of 310 ACL anthology papers. The original dataset is consisted of 8736 citations out of which 829 are labelled as positive whereas, 280 are tagged with negative sentiment. Because the basic unit of interest is the citation sentence and its relevant location in the cited research paper. I collected the PDF versions of the papers to which the citation sentences belong, extracted the metadata (XML Format) and sections/location in which the citation sentence appears. For the sake of this experimental study, I primarily focused on sentiment bearing citation sentences (p – positive, n – negative) and also randomly selected objective sentences. I stretched the scope of citation sentences and manually added the citation context for each of the citation sentences. Since the context is consisted of more than one sentiment per citation text, I want to pragmatically analyze the drift in the author's sentiment and identify the real intention towards the cited work within a contextual window. I added up to 2 posterior contextual sentences and 2 sentences before the citation sentence. I annotated each contextual sentence with a three-class scheme i.e., positive 'p', negative 'n' and neutral/objective 'o' and relevancy with the citation sentence. The underlying corpus is consisted of approximately 2109 citation contexts which correspond approximately to 24% of the original dataset. The objective was to study the behavior and impact of citation context in which a citation appears on the sentiment polarity of citation sentences. In the existing dataset, there is no information in regards to the rhetorical structure where the citation sentence exists. So, the existing dataset was extended for extracting the location of each sentence in the research paper. The next task was the categorization of sections based on a four-section structure: Introduction, Method, Result and Discussion. The citation contexts were also annotated with the paper sections ('Introduction', 'Literature Review', 'Methodology', 'Results', 'Discussion', 'Conclusion' and others) in which they appeared. The section labels and names allowed the alignment of research papers in our dataset. For this sake, the section titles in which the relevant citation sentences appeared were analyzed and tagged based on the regular expressions corresponding to each section type considering the potential variants of the section titles. More than 98% of the citation sentences in the corpus pertain to these four sections but not necessarily in the same sequence. For this particular study, I didn't consider the order of sections in a research article. The citation sentence, its corresponding contextual sentences along with sentiment polarity, section appearance and relevancy are extracted and saved in the Oracle database. An example annotation is depicted in the figure below with shows citation context, its class label along with XML file for reference information.

3.11.3 Evaluation Metrics

To evaluate the performance of the proposed technique in identifying the sentiment of the author towards a citation text, the standard performance metrics of precision, recall, and accuracy are employed, which are commonly used in the evaluation of various machine learning schemes. The trade-off between these metrics depends upon the systematic application to the problem in hand. For the citation text corpus, these evaluation metrics are calculated using the formulas presented below: The percentage of the correctly classified test set (TP + TN) to the total number of instances (TP + FP + TN + FN) is defined as the accuracy or the recognition rate of the classification problem. Accuracy determines how well a classifier recognizes the instances of various class labels and is computed with the formula:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} * 100 \tag{3.17}$$

Precision or positive predicted value is the percentage of correctly classified predicted instances. In other words, this is the percentage of truly predicted (TP)

Turney (2008) has recently proposed a simpler SVM-based algorithm for analogical classification called PairClass. While it does not adopt a set-based or distributional model of relational similarity. We have noted above that PairClass implicitly uses a feature representation which does not produce compatible results.

citationText	Turney (2008) has recently proposed a simpler SVM-based algorithm for analogical classification called PairClass.	P
postSentence1	While it does not adopt a set-based or distributional model of relational similarity.	N
postSentence2	We have noted above that PairClass implicitly uses a feature representation which does not produce compatible results.	N


```
- <annotatedPaper id="1">
  - <paper title="Using lexical and relational similarity to classify semantic relations " year="2009" authors="Diarmuid O Seaghdha, Ann Copestake">
    - <citationContext>
      - <citationText>
          Turney (2008) has recently proposed a simpler SVM-based algorithm for analogical classification called PairClass.
          <section> Literature Review </section>
          <cite polarity="p">Turney (2008)</cite>
        </citationText>
      - <postSentence1>
          While it does not adopt a set-based or distributional model of relational similarity.
          <cite polarity="n">Turney (2008)</cite>
        </postSentence1>
      - <postSentence2>
          We have noted above that PairClass implicitly uses a feature representation which does not produce compatible results.
          <cite polarity="n">Turney (2008)</cite>
        </postSentence2>
      </citationContext>
    </paper>
  </annotatedPaper>
```


Figure 3.5: An example of citation context containing explicit and implicit sentences.

instances to the total number of positively predicted class labels. It is the measure of exactness and a high percentage of precision demonstrates more relevant results produced by the technique as compared to irrelevant results. Precision is measured using the following formula:

$$Precision = \frac{TP}{TP + FP} * 100 \tag{3.18}$$

Recall or sensitivity is measured as the number of correctly classified instances of a class (TP) divided by the sum of TP (number of correctly classified instances) and FP (number of incorrectly classified instances means that tuples belong to this class but incorrectly classified to the other class). The recall is the measure of the completeness of the results and a high value of sensitivity demonstrates that most of the results produced by the technique are relevant. It is measured by the following formula:

$$Recall = \frac{TP}{TP + FP} * 100 \tag{3.19}$$

F-Measure is calculated by taking the harmonic (weighted) mean of precision (exactness) and recall (sensitivity or completeness). There exists a direct relationship between the values of precision and recall and the value of F-Measure. So, if precision and recall measures have high values then the value of F-measure will also be high and a higher F-Measure value means better classification accuracy. F-measure is widely used in the domain of text classification for performance evaluation and it is well-suited to our work because of the high unbalanced nature of the Citation Sentiment dataset. Among the four, evaluation metrics accuracy evaluates the overall correctness of the classification results whereas the other measures deal with evaluating the correctness for each class. It is calculated with the help of following formula:

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.20}$$

Chapter 4

Results and Discussion

4.1 Citation Sentiment Analysis (RQ1)

In this section I will be presenting the findings against RQ1 of this study which is related to the impact of higher-order adjective, adverb and their combination n-gram word phrases on the citation sentiment classification and this belongs to the first research problem described in Chapter 1. To answer this question, I have used a pre-labelled and manually classified corpus for the experiments to serve as a baseline for comparing and evaluating the results of the technique. Moreover, I have also labelled and classified the data set consisting of the citation text using two commercial sentiment analysis tools SEMANTRIA and THEYSAY. I compared the results of classification by our technique with the manually classified text and with the commercial tools. The figure 4.1, 4.2 and 4.3 represent the percentage of positive, negative and neutral citations classified by manual and commercial tools. The results of the classification performed by commercial tools are different which establishes that these tools are implemented based on different internal algorithms. Different percentages of positive, negative and neutrality describe that both the tools handle the negativity and neutrality in different ways. According to SEMANTRIA among these 8730 citation texts, 2264 are in positive sentiment class, 1266 are with negative sentiment orientation and the remaining is

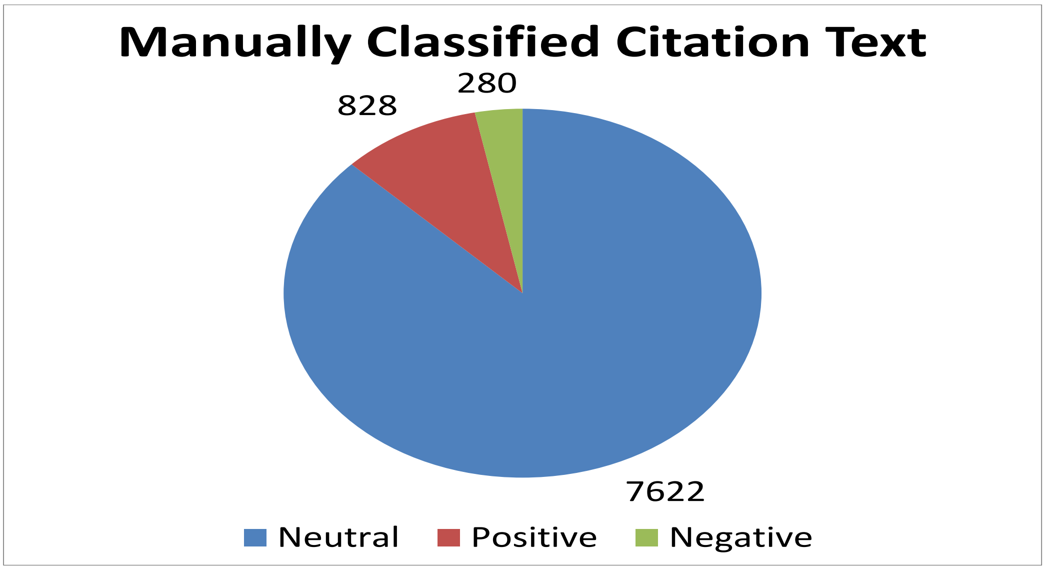

FIGURE 4.1: Class label percentage of manually classified citations text

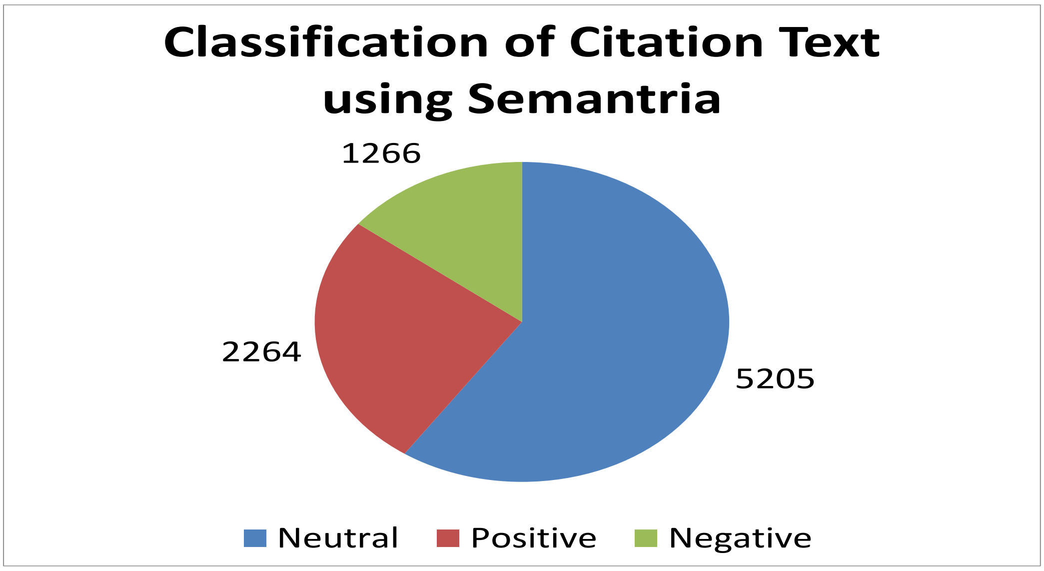

FIGURE 4.2: Percentage of class labels produced by SEMANTRIA

labeled as neutral. Whereas, according to THEYSAY this distribution is as 4802 positive, 2130 negative and 1799 as neutral. In the manually labeled data set the majority of the citations labelled as neutral or objective, i.e., 87%, whereas the percentage of neutrality or objectivity is quite less in the commercial tools which are 59.5% by SEMANTRIA and 20.5% by THEYSAY, respectively. The classification performed by manual annotation is highly skewed towards the objectivity and the same pattern is depicted by SEMANTRIA. This shows that SEMANTRIA depicts a clear tendency of assigning objective labels to the citation dataset

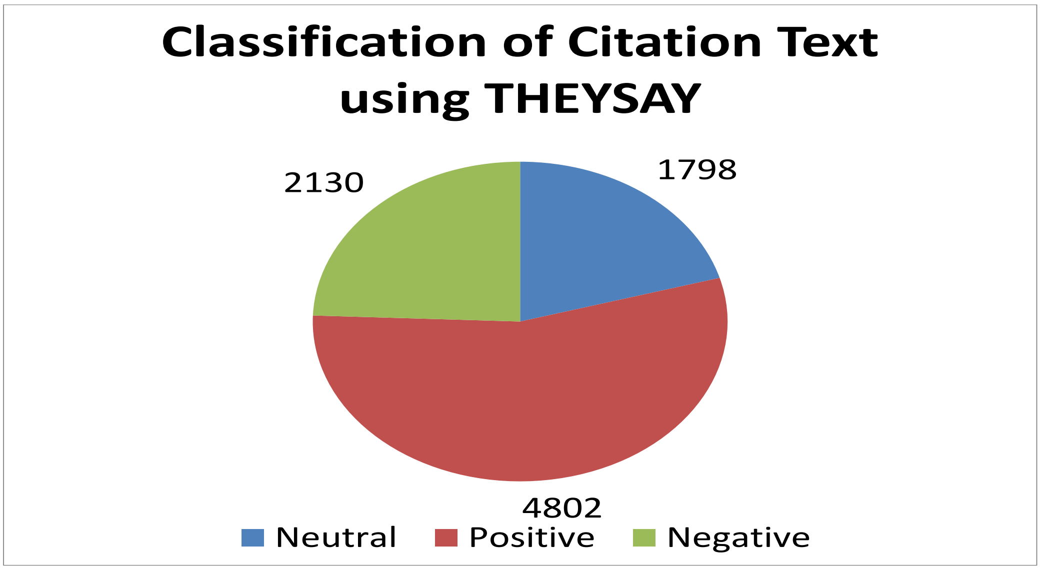

FIGURE 4.3: Percentage of class labels produced by THEYSAY

as compared to the THEYSAY which demonstrates a biased behavior towards the positive class. When comparing the tendency of class labels produced by commercial tools with the baseline manually annotated dataset, it is established that results produced by SEMANTRIA are closer to the benchmark so far as all the class labels are concerned. The presence of adjectives and adverbs in a citation text is utilized to classify it as positive, negative and neutral. In experimentation, I extracted the bag of words consisting of adjectives and adverbs based on n-grams. For this purpose, I have not only considered unigram (n = 1) and bigrams (n = 2) but have also exploited the use of trigrams (n = 3) and pentagrams (n = 5). The most specific reason for using the higher-order n-grams is because most of the state-of-the-art approaches claim that they can play a more significant role than lower order n-gram in sentiment detection. Therefore, in this study, I have investigated how different high order pair of words, behaves in determining the sentiment orientation on the citation dataset. I have not only considered the adjectives and adverbs individually in determining the sentiment polarity of the citation text but have also evaluated them on their different combinations. The usage of the high order n-grams as input feature resulted in high dimensional feature space (19,648 dimensions of unigram, 55,796 dimensions of trigrams and 87,599 dimensions of pentagrams) in our scenario. The extracted bag-of-words are sorted as per the

Table 4.1: Top 20 Frequent Adjective Terms.

Most Frequent Adjective Unigrams (n=1)	High, good, successful, effective, popular, efficient, important, interesting, robust, difficult, considerable, easy, competitive, suitable, poor, central, appealing, straightforward, satisfactory, dominant
Most Frequent Adjective Bigrams (n=2)	such as, statistical machine, natural language, good performance, significant improvements, human judgments, minimum error, maximum entropy, automatic evaluation, the same, human evaluation, more accurate, in recent, current state, Recent work, more sophisticated, very successful, be effective, most notable
Most Frequent Adjective Trigrams (n=3)	such as the, minimum error rate, statistical machine translation, with human judgments, in recent years, phrase based statistical, from many heterogeneous, In recent years, Introduction In recent, the past few, to be effective, past few years, successful in recent, current state of, greedy search algorithm,various NLP tasks, techniques such as
Most Frequent Adjective Pentagrams (n=5)	proven increasingly successful in recent, increasingly successful in recent years, major developments in statistical approaches, current state of the art, notable examples of unsupervised polarity, received much attention in recent, many advances in recent years, such as part of speech, past five years important research, Introduction In recent years statistical

Table 4.2: Precision, Recall and Accuracy of classified citations.

			Manual		SEMANTRIA				THEYSAY
Feature	**P**	**R**	**F-score**	**P**	**R**	**F-score**	**P**	**R**	**F-score**
Adj. (1-g)	70.54	40	55.27	75.5	29.17	52.34	44.45	80	62.23
Adj. + Adv. (1-g)	75.54	55.56	65.55	94.11	55.56	74.84	53.35	78.17	65.76
Adj. (2-g)	75	50	62.5	70	54.17	62.09	55.56	76.92	66.24
Adj. + Adv. (2-g)	88.89	80	84.45	85.5	92.30	88.9	58.06	85.71	71.89
Adj. (3-g)	65.55	50	57.78	96	58.33	77.17	70.54	80	75.27
Adj. + Adv. (3-g)	88.89	72.72	80.81	96	68.57	82.29	59.38	82.60	70.99
Adj. (5-g)	75.54	60	67.77	94.11	66.67	80.39	54.17	81.26	67.72
Adj. + Adv. (5-g)	90	81.82	85.91	96.15	96.15	96.15	55.88	82.60	69.24

frequency of their occurrence in the dataset to select the most important ones. The extensive reason for this was to identify the frequent terms in the scientific literature for sentiment detection. Table 4.1 shows the list of most frequent unigram, bigram, trigram, and pentagram adjectives and adverb combinations that infer the opinion. In figure 4.4, 4.5, 4.6 and 4.7 I plotted the frequency of the terms related to the citation text including their variations.

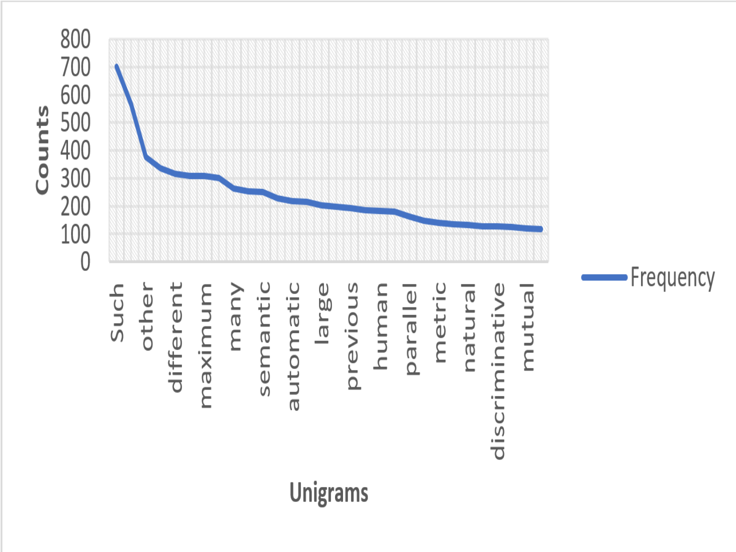

Figure 4.4: Frequency Graph of Adjective Unigrams

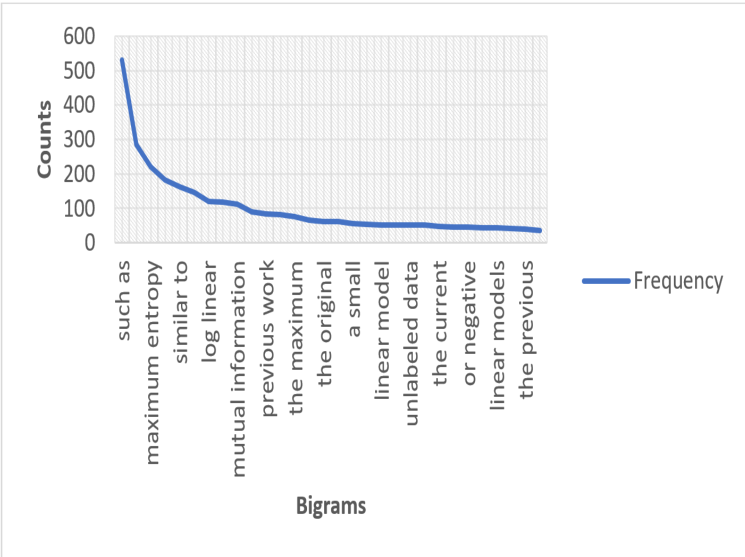

Figure 4.5: Frequency Graph of Adjective Bigrams

In this section, I present the results of citation sentiment classification by exploiting adjective, adverb and their combinations by means of high order n-grams at the document level. I have not only compared the results obtained by our approach with the manually annotated corpora, but also juxtaposed the results with classified citations using two commercial tools. In Table 4.2, the values of evaluation metrics (precision, recall and recognition rate) are presented and a comparison is made with manual annotation and classification results against annotation performed by the commercial tools. It can be analyzed that the value

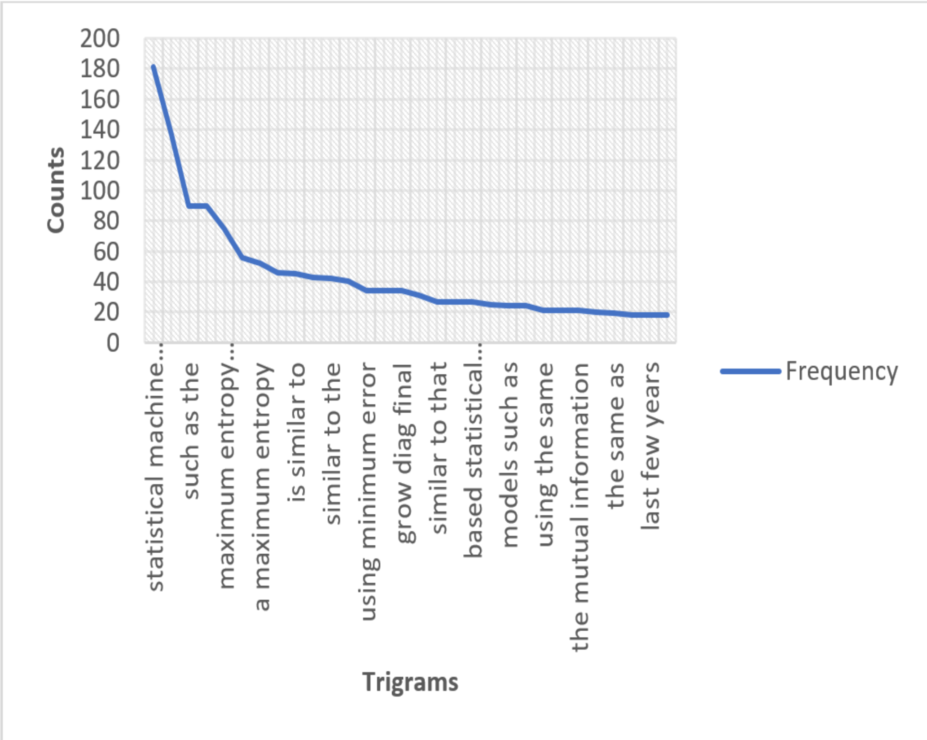

Figure 4.6: Frequency Graph of Adjective Trigrams

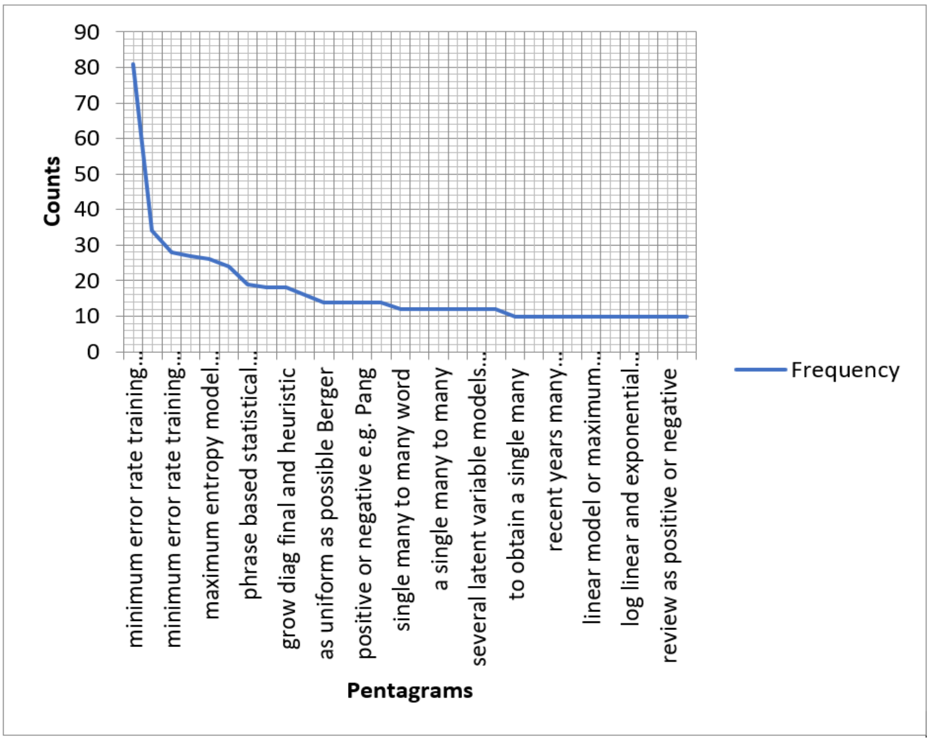

Figure 4.7: Frequency Graph of Adjective Pentagrams

of accuracy obtained using higher-order n-grams is better than the value obtained using lower-order n-grams. Hence, when I carried out the analysis using "single word i.e., unigram", "double word i.e., bigram", "triple word i.e., trigram", "five words i.e., pentagram", adjectives, adverbs and their combinations. The accuracy value obtained using adjective and adverb combinations is comparatively better than that obtained using the adjective and adverbs separately. Trigram and pentagram outperform the lower-order n-grams which establishe the use of higher-order n-grams as a suitable feature for the citation corpus. Let's now discuss the accuracy of the findings of classification results. When I have selected n-grams as features then semantic information is partially lost or neglected more specifically in the case of BOW i.e., unigrams. The outcomes of the proposed study show an accuracy of above 80% against the manually annotated corpus using higher-order n-gram adjective and adverb combinations, as explained in Figure 4.8. It can be observed from the results that with an increase in the value of n, the classification accuracy has increased more specifically when adjectives and adverbs are used in a combinatory fashion. This further affirms that higher-order n-gram adjective and adverb combinations are more precise and deterministic expressions than the lower-order n-grams. With an accuracy of above, 90% for adjective and adverb combinations SEMANTRIA has produced the most precise predictions, following the same trend line as depicted against the manual baseline. This shows that these features (high order n-gram adjective and adverb combinations) produce better results as compared to unigrams and bigrams. The analysis of the results has revealed that the accuracy of the lower order n-grams ($n = 1$) remained the same for individual features (adjectives and adverbs) and both in a combinatorial fashion. The results also asserted better accuracy for average positive citation texts as compared to the negative and neutral ones. Furthermore, the objective citations were misclassified as either positive or negative. The possible reason for this is that it is difficult to predict neutrality and negation. The analysis of the results has also stated that the use of higher-order n-grams might solve the problem of compositionality (understanding a complex expression through the meanings of its constituent expressions). It is again worth mentioning, that the accuracy de-

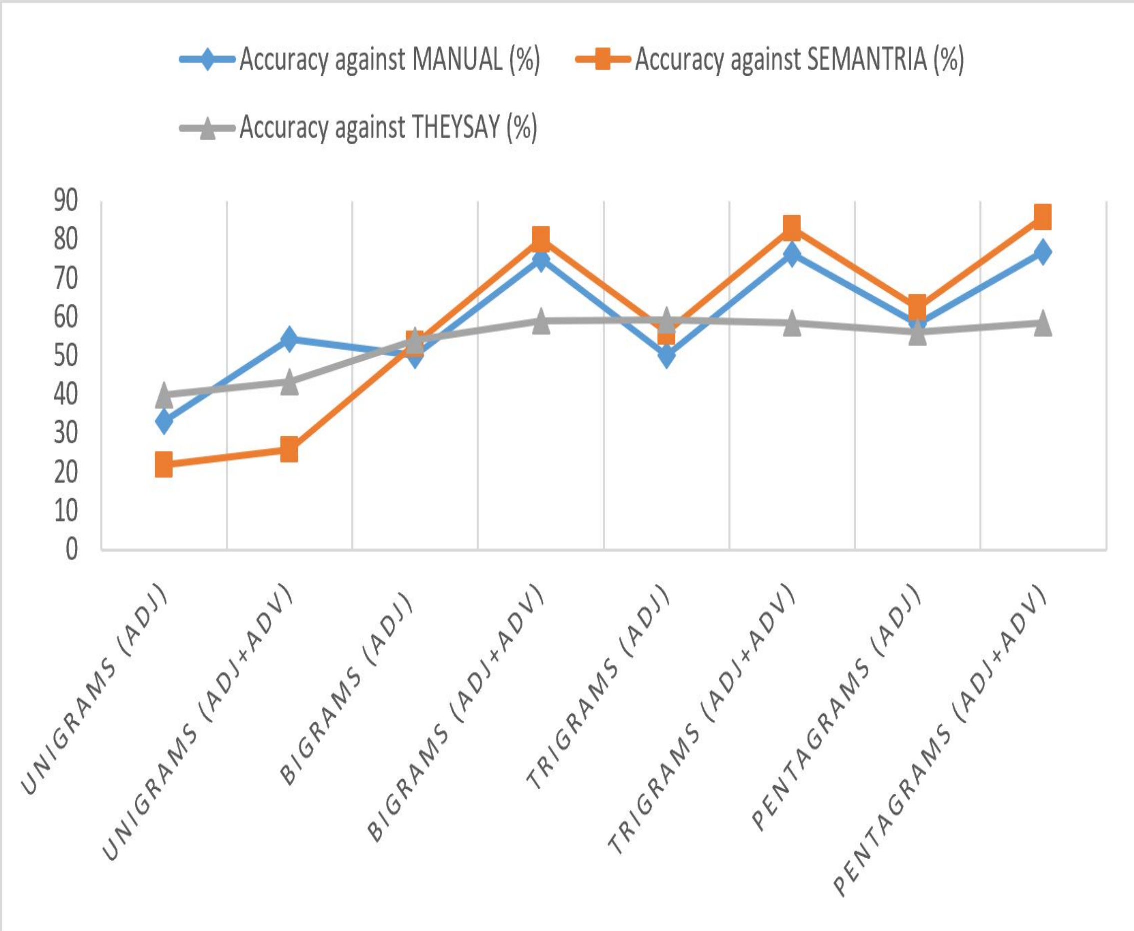

Figure 4.8: Comparison of accuracy against baseline

picted against THEYSAY is persistently lower as compared to the SEMANTRIA and manual annotation of all the features. This demonstration is mainly because of the biased predictive behaviour of THEYSAY towards the positive class. However, it is also evident for the accuracy trend line of THEYSAY that adjective and adverb combination has obtained a high accuracy rate as compared to the isolated features. This also confirms and strengthens the accuracy of classification results. Though SEMANTRIA may not be considered as the best tool for the sentiment analysis, however, it has shown a persistent behaviour for different datasets i.e., healthcare survey data (Georgiou et al. 2015) and multilingual datasets (Tandon and Jain 2012). THEYSAY has performed incorrect classifications for large explanatory sentences. As the average sentence size of the experimental dataset was 35 words which caused THEYSAY to make the majority of false predictions. In Figure 4.9, for adjective and adverb combinations and high order n-grams, the proposed technique has achieved an average precision of about 90% against manual classification and 96.15% against SEMANTRIA, respectively. Thus, the effectiveness of the proposed technique is in the considerable range when comparing against manual classification and annotation results against SEMANTRIA. It is also worth mentioning that our proposed technique has acquired a substantial improvement in precision as compared to recall. The precision and recall against THEYSAY are consistently low for all the features. The high precision and recall values can be observed from Figure 4.9 4.10 for both individual and adjective and adverb combinations against SEMANTRIA, which further ensures the reliability of the said tool. The combinations of adjectives and adverbs based on high order n-grams are a better choice in predicting the document-level sentiment for larger sentences. Further, to examine the computational load, I have examined the impact of the size of citation corpora on the performance of the proposed system. The results are presented in the Figure 4.11. As it can be depicted from the graph that as the instances in the sample size are increased, there is an improvement in the performance of the system. However, at a certain point in time, the improvement may not be achieved by merely increasing the size of the training dataset when the size of the corpus is large enough.

4.1.1 Performance Evaluation

In this section, I present a comparative analysis of the results obtained using the proposed approach to that of the other literature using citation corpus and n-gram approaches. The comparative analysis of the results is presented in the Table 4.3. [1] used n-grams of different lengths indicating that unigram, bigram and trigram performed better when used in combination as compared to the unigram and unigram plus bigrams. Our work is closely related to the work done by [1, 9] and is based on the hypothesis stated by them "using higher-order n-grams might prove to be useful in sentiment detection". The same has been endorsed by our proposed approach that higher-order n-gram adjective and adverb combinations give better results as compared to lower-order n-grams. Because lower-order n-grams

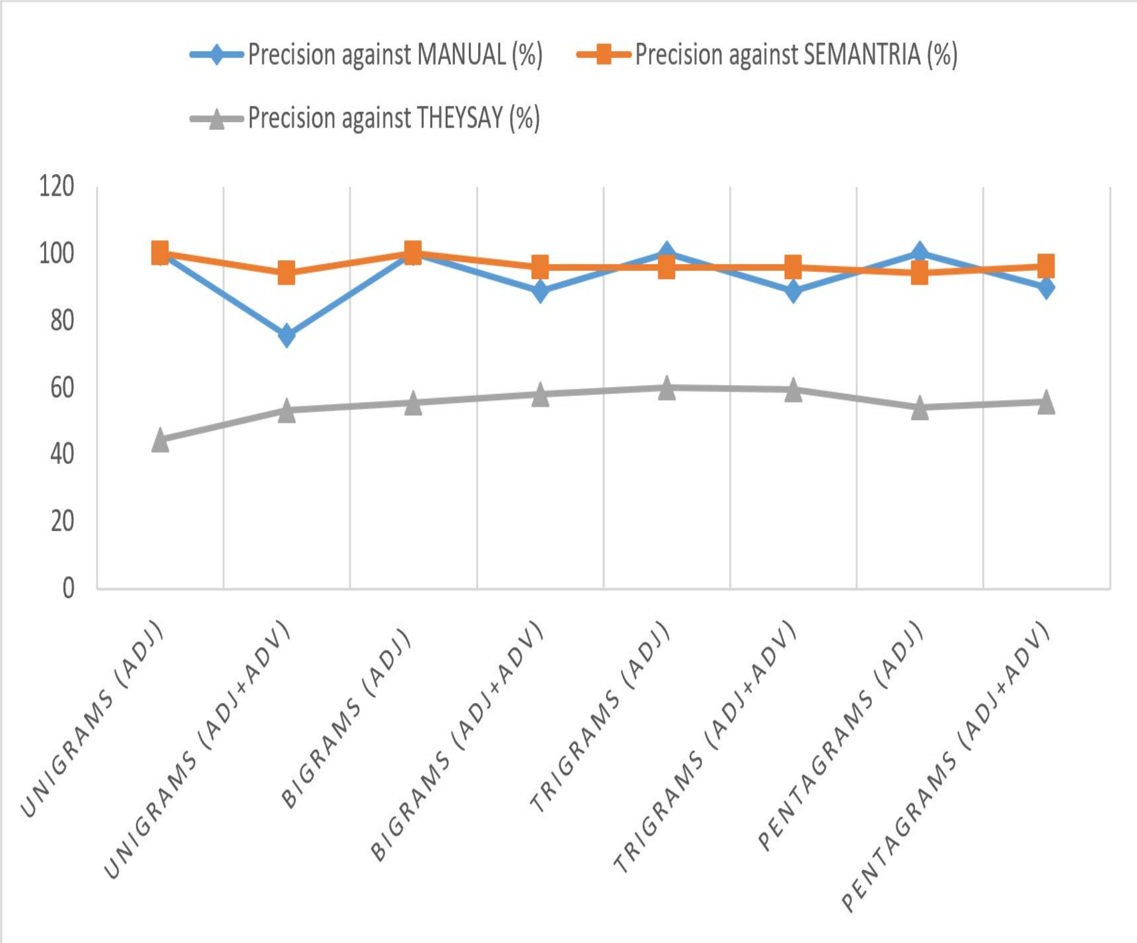

FIGURE 4.9: Comparison of precision against baseline

are not able to capture the longer-range dependencies which might hinder the classification accuracy. Another insight from both the studies is that both have been carried on the same annotated citation corpus. [7] used n-gram, verbs, adjectives, and their combinations to generate the structured tabular summary of the cited work. Their results depict that a combination of adjectives, adverbs and bi-grams give an average accuracy of 68.54%. Their study was carried out on a relatively small corpus consisting of only 500 citation contexts. Our proposed approach contributes that adjectives and adverbs combined with higher-order n-gram give better results with an accuracy of 76.9%, on a manually annotated corpora consisting of 8736 citations as shown in Table 4.3. [23] conducted the study to identify the sentiment polarity from the citation context on a dataset consisting of 4182 citations extracted from 285 randomly selected clinical trial papers. The possible similarity between our model and theirs is that both the techniques are

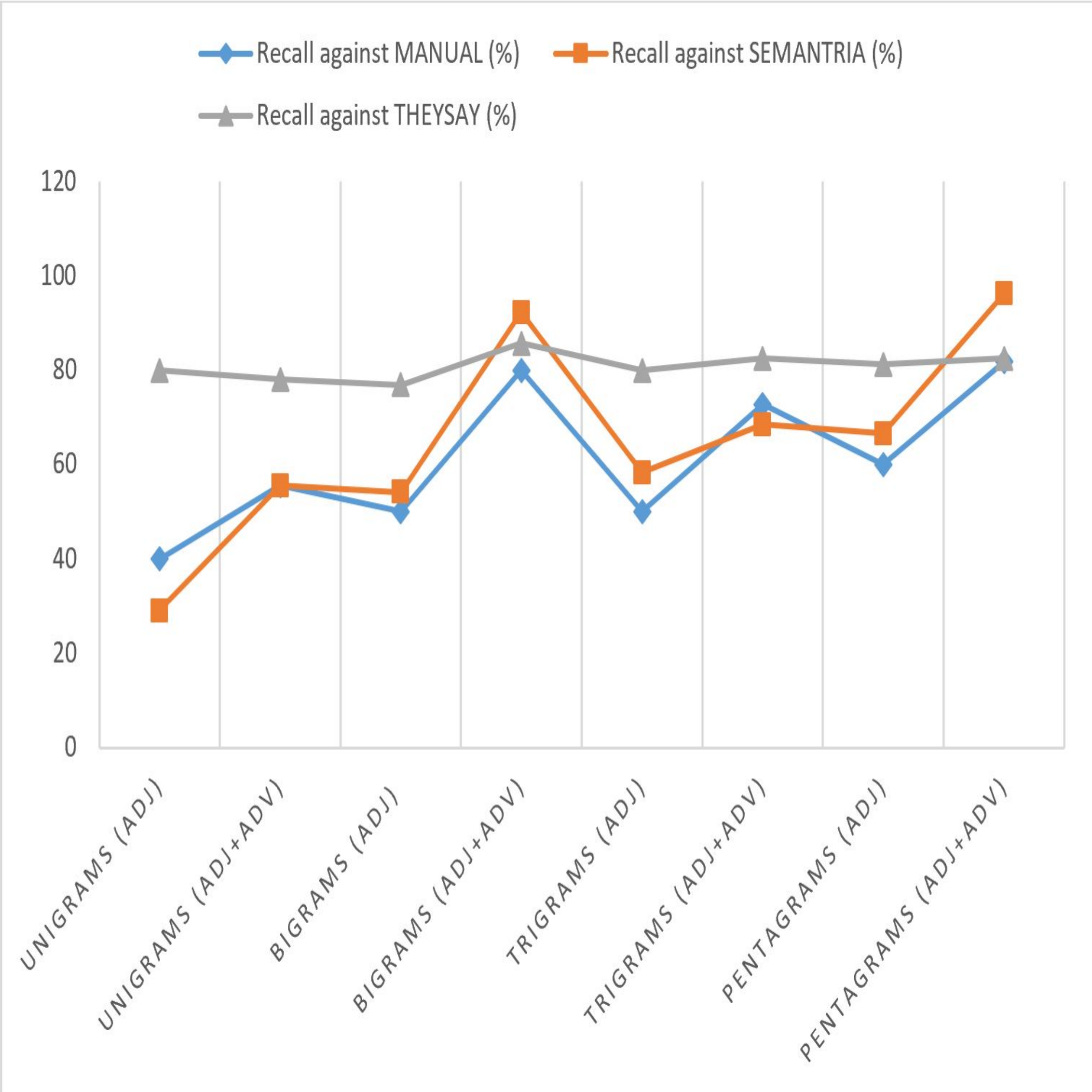

FIGURE 4.10: Comparison of Recall against baseline

unsupervised. [5] have presented an SVM based text categorization technique for the author's sentiment classification. They have carried out their classification experiments using a corpus consisting of 2,665 citation sentences. As per the findings of both the studies lower-order n-grams give better results as compared to the higher-order n-grams. Both the studies have been performed on a dataset extracted from biomedical articles. As the data domain being explored affects the classification results. The results cannot be generalized because I performed our experiments on a different dataset belonging to the computer science domain. The literature in the computer science domain is mostly about the theoretical models, frameworks, algorithms and application systems. Whereas the researchers in the

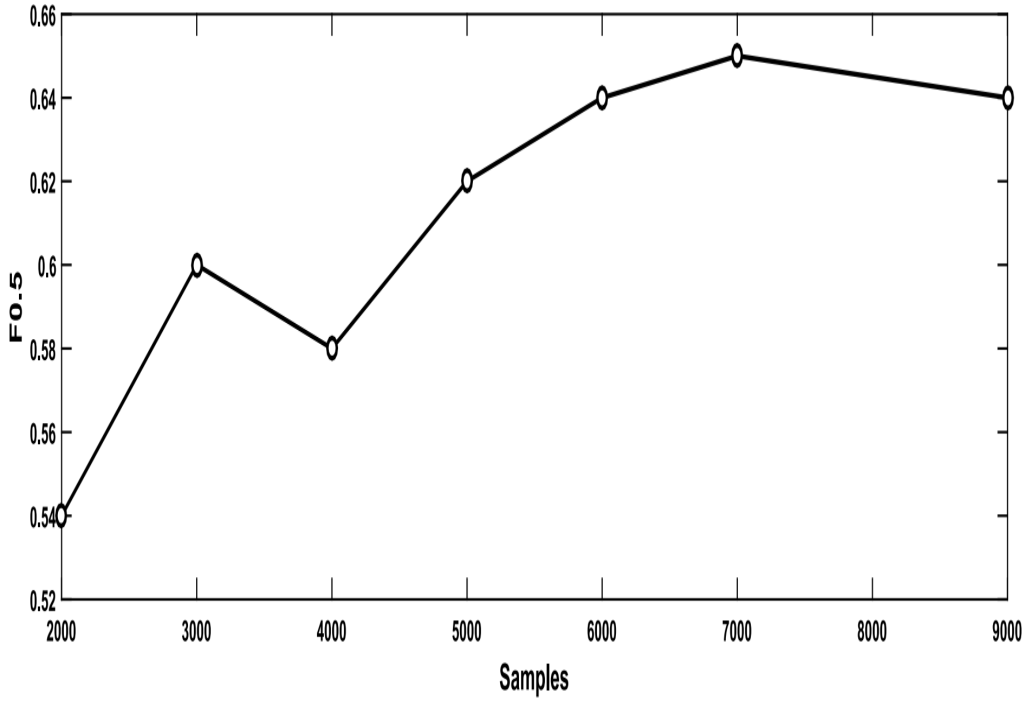

FIGURE 4.11: Impact of varied size of citation corpus on F0.5 Measure

biomedical domain express their sentiments indicating whether their work confirms, supports or agrees with the cited literature. Another significant difference between our proposed approach and that used by [5] is that they are using a supervised approach for the classification whereas our approach is unsupervised and domain-independent. So this scenario ascertains a need for a more comprehensive and detailed corpus encompassing tuples from multiple domains for the generalization of results.

The performance of the model is best in predicting positive citation sentiments, followed by negatively cited citations and then the neutral citation sentences. Most of the errors are due to neutral citation being classified as positive or negative because of the biasness of the learning algorithms towards the neutral class. To remove the sensitivity of the classifier towards neutral class, a subset of neutral citations was randomly selected using a predefined ratio of positive and negative citations as compared to neutral examples. The proportion of neutral class examples (87.3%) to negative (3.2%) and positive (9.5%) class instances in the underlying dataset is 1: 3: 9, for the experiments the ratios of 1: 1 and 1: 2 in the

TABLE 4.3: Comparative Analysis of Results with Literature using Citation Corpus and N-gram Approach.

	Athar (2011)	**Tandon & Jain (2012)**	**Kim & Thoma (2015)**	**Athar & Teufel (2015)**	**Xu et al. (2015)**	**Proposed Approach**
Features Used	N-grams (1-3)	N-grams, Verbs, Adjectives, Verb & Adjective Combinations	N-grams (unigrams & bigrams)	N-grams of length 1 to 3	N-grams (2-words to the right of negations)	(1-grams):Adj, Adj+Adv (2-grams):Adj, Adj+Adv (3-grams):Adj, Adj+Adv (5-grams):Adj, Adj+Adv
Results Obtained	N-grams (1-3 grams): macro-F: 0.597, micro-F: 0.862	Adj.: 65.54, Verb: 66.30 Adj.+ Verb: 67.48 Adj.+Verb+ Bigram: 68.54	Unigram: 84% Unigram+ Bigram: 86%	F Macro: 0.731 F Micro: 0.871	Accuracy: 0.853	(1-grams):Adj: 0.33, Adj+Adv: 0.543 (2-grams):Adj: 0.50, Adj+Adv: 0.75 (3-grams):Adj: 0.50, Adj+Adv: 0.763 (5-grams):Adj: 0.583, Adj+Adv: 0.769
Corpus Used	Manually annotated 8736 citations from 310 research papers	500 labeled citation contexts	2,665 CON sentences from online biomedical articles	1,741 annotated citations	Number of Citations: 4182	We used the same dataset as of Athar

under sampling. One of the implications of using the under-sampling technique is some useful information may be lost but this is the inherent drawback with the approach. The neutral class is hard to predict and in future, I intend to use a larger citation corpus to train the model for eliminating the bias towards neutral class and other under sampling techniques like Ensemble Methods and hybrid approach.

4.2 Aspect based Citation Sentiment Analysis

4.2.1 Frequent Aspects Detection

The first contribution of our experimental work is to discover the frequent aspects in the citations which authors usually discuss while citing a paper. From a particular citation text, I extract the nouns phrases on the basis of the criteria mentioned in the methodology section and obtain a sentiment score for each of the cited aspects. I present here only those aspects which have been evaluated on some heuristic rules (1) for which opinion words have been used to describe them and (2) the aspect must be discussed the most in the citation sentences. Table 4.4 presents the most frequent aspects from both the corpora and sentiment polarity distribution over aspects is specified in Table 4.5. It is worth mentioning here that

Table 4.4: Discovered Aspects from Comparative Domain Corpora.

CS Corpus	**BI Corpus**
Discovered Aspects	**Discovered Aspects**
Model; performance; method; approach; Algorithm; system; results; usage; technique; accuracy; parser; Metrics; corpus; measure; feature; corpora; data;correlation; research; work	Study; Results; Findings; clinical trial; performance; Data; technique; Correlation; efficacy; Association; usage; Regression; research; hypothesis; success rate; failure rate; response rate; work; analysis
Technical Verb	**General Verb**
Statistical, syntactic, semantic, lexical, probabilistic, unsupervised, discriminative	Such, good, recent, efficient, many, successful, significant

Table 4.5: Sentiment polarity distribution over aspects.

Sentiment	**No. of Aspects**	**Aspect Percentage**
Positive	132	65%
Negative	71	35%
Total	203	100%

the citation corpus is heavily unbalanced, both in terms of the aspect distribution as well as in terms of sentiment polarity mentions. The results revealed that from all the discovered aspects, the dominant sentiment category is 'positive'. Figure 4.12 illustrates the number of labelled citations with each aspect along with its positive and negative sentiment polarity. It becomes obvious from Figure 4.12 that authors usually prefer to discuss the 'methodology', 'performance', 'corpus', 'study', 'measure', 'usage' and 'result' aspects. The ranked list of significant verbs used to describe different aspect categories is shown in Table 4.4. Moreover, a fascinating fact is revealed that there are two types of verb categories one that is used for general purpose and the second one is mostly consisted of technical terms.

For each aspect category, there are certain words which when described with an opinion word would rather depict the author's sentiment about the aspect category. Considering the citation sentence '. . . Our outcome is worthwhile because it is consistent with prior results. . . ', in which the word "worthwhile" describes the aspect word "outcome" which indicates the aspect category "Findings". The aspect cue words are mapped into the potential aspect categories and the most frequent one is assigned as being the leader of the aspect category. To consider extracted noun phrases to be potential aspects, the proposed model relies on the chi-square scores. The pre-conditions for applying the Chi-square test was met considering: (1) two-categorical variables (Citations containing unigram term and citations not containing unigram term) (2) two or more classes/categories (No. of positive, No. of negative and No. of neutral citations) for each of the variable under consideration (3) Observations are independent i.e., there was no pairing between the categorical variables in both pre-test and post-test observations (4) underlying dataset is consistent in regards to expected distribution in the form of aspect polarity distribution as positive or negative. The various aspect categories along with aspect cue words extracted from the citation corpus are presented in Table 4.6. I checked the results for ensuring the accuracy of the aspect detection with the help of manual inspection. The most frequent aspect category is with the title "technique" along with its cue phrases (model, method, approach, process, etc.) with over 571 subjective citations, followed by the "performance" aspect category (417) and corpus (267). While the occurrence frequency of aspects like "feature" and "parser" is less than 100 citations. Low ranking aspects does not mean that these do not reflect the citation intention, however, it indicates that the authors tend to be more attentive towards the high ranked aspects. Another observation is that domain-specific aspects are less frequent as compared to generic aspects. It can also be noticed that some frequent aspects i.e., "findings", "performance" and "study" are homogeneous in both the corpora. Analysing the results at the most fine-grained level, the probability to detect positive sentiment at the aspect level is higher than the negative mention.

I measure the performance of the proposed aspect identification technique by making a comparison of the results produced by the system with the labelled aspects. It can be seen from the Table 4.7 that the proposed approach produced encouraging results with an overall accuracy with bigram (70.75%), trigram (79.25%) and bigram trigram in combination (81.39%) respectively when compared with the manually tagged aspects. The bigram, trigram and pentagram rule patterns

Table 4.6: Citation aspects along with associated aspect clue words.

Aspect Category	**Aspect Clue Words**	**Conf. Interval**
Findings	Conclusion, outcome, result, consequence, resultant, answer	90% (>2.706)
Corpus	Corpora, dataset, data	90% (>2.706)
Technique	Model, method, approach, metrics, modelling, process, procedure, algorithm, system, framework, simulation, pattern, function	95% (>3.841)
Study	Research, work, report, paper	90% (>2.706)
Measure	Regression, correlation, association, efficacy, coefficient, criterion	90% (>2.706)
Performance	Accuracy, recall, precision, execution, functioning, success rate, failure rate, response rate, log-likelihood ratio	95% (>3.841)

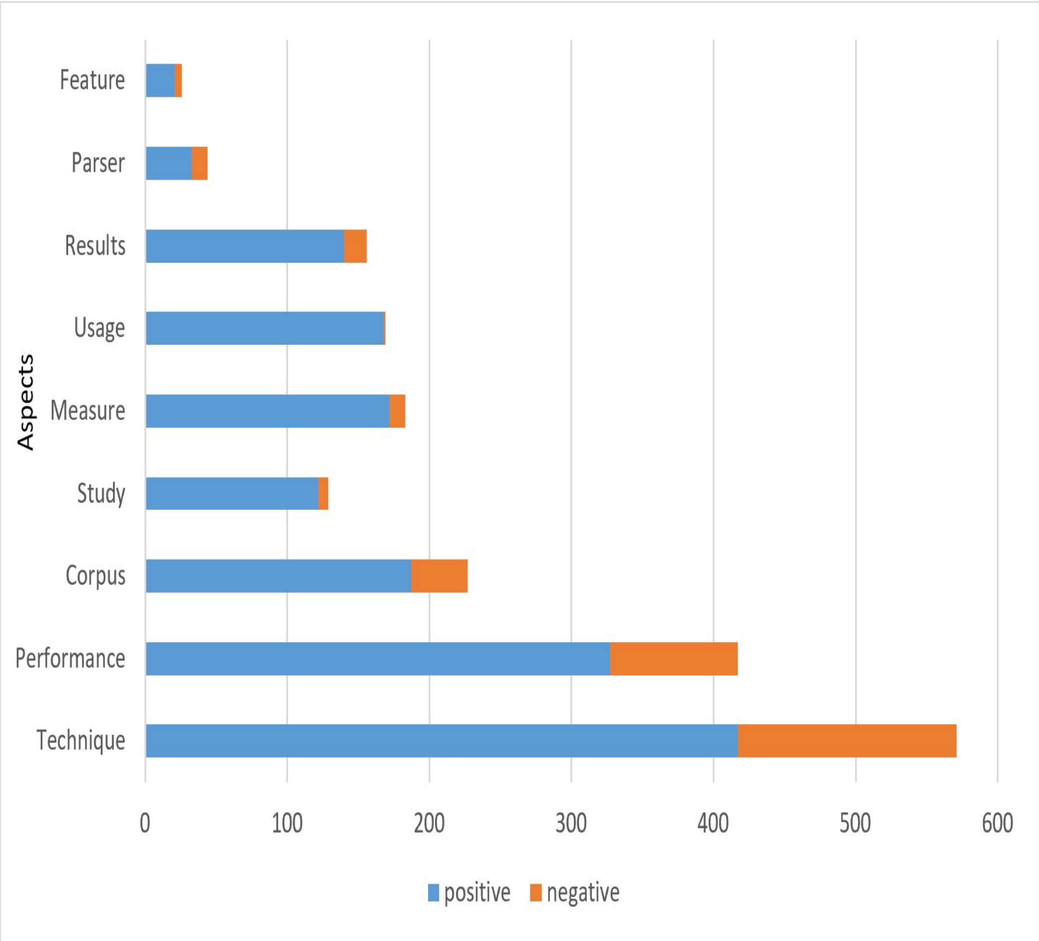

Figure 4.12: Annotated Aspects along with Sentiment Polarity

Table 4.7: Aspect identification.

Method	Citations	Manual tagging	Proposed model Output	Accuracy	Overall accuracy
Bigram	Positive	500	364	72.8	70.74
	Negative	313	215	68.69	
Trigram	Positive	500	413	82.16	79.25
	Negative	313	239	76.35	
Bigram + Trigram	Positive	500	421	84.2	81.39
	Negative	313	246	78.59	

will not only cover the single word aspects, multiple word aspects and the proper nouns. The results indicate that utilizing the pattern knowledge, synonyms and chi-square test for determining the participation level of aspect produced significant results in the process of aspect identification. The manifestation of different types of linguistic patterns will leverage the benefit of minimum unwanted aspects. However, the results also indicate that there are certain aspects which have been tagged by the experts but not identified by the proposed approach which I attribute towards the presence of infrequent patterns, low participation level and lack of synonyms for the technical noun phrases.

4.2.2 Citation sentiment classification results (RQ2, RQ3 and RQ5)

In this experiment, I test the aspect based citation sentiment classification using three methods ('N-gram Before', 'N-gram After', 'N-gram Around') which is our RQ2 and third research problem described in Chapter 1. The positive and negative polarity of nouns, adjectives, verbs and adverbs is identified that is located near to the linguistic expressions that represent a given aspect. For the evaluation of the method's accuracy, it is necessary to analyse their efficiency in the citation sentiment classification process. For this sake, I explored the different values of the POS n-gram parameter (ranging from 2 to 5) to discover the best setup covering adjectives, verbs, adverbs and noun phrases which is RQ5 of our research study. I run our experiments for citation polarity classification by removing the

TABLE 4.8: Citation sentiment classification obtained with the 'N-Gram After' method.

	N=2				N=3				N=5			
	P	**R**	**A**	**F**	**P**	**R**	**A**	**F**	**P**	**R**	**A**	**F**
SVM	68.2	68.2	68.22	56.3	70.9	70.3	70.34	59.4	75.4	74.2	74.15	63.9
Naïve Bayes	67.9	67.9	67.87	54.9	70.84	70.6	70.56	58.9	74.6	73.9	73.89	67.89
MaxEnt	66.9	66.9	66.89	53.6	61.8	69.9	69.89	59.1	73.9	73.6	73.6	65
J48	67.45	67.12	66.95	53.7	70.42	70.12	71.2	57.9	72.6	72.6	72.63	61.1
Random Forest	69.8	69.8	69.79	57.4	73.9	72.1	74.03	66.1	79.5	79.9	79.9	78.1
Average	**68.05**	**67.98**	**67.94**	**55.18**	**69.57**	**70.6**	**71.2**	**60.28**	**75.2**	**74.84**	**74.83**	**67.2**

citations belonging to the "neutral" class and considered only the "positive" and "negative" citations. The results are reported in terms of precision (P), recall (R), accuracy (A) and f-measure (F). For the classification experiments, those aspects are considered which are identified by applying the proposed aspect detection technique. This resulted in a reduced number of unwanted and irrelevant aspects ensuing in a considerable reduction in feature space. The results presented in Table 4.8, Table 4.9, Table 4.10 illustrate the average citation sentiment classification accuracies using different ML classifiers for comparative domain corpora which is related to RQ3. The results corresponding to 'N-gram after' method are outlined in Table 4.8. It can be observed from the results that for the citation sentiment classification task the best average success rate is obtained with the higher-order n-gram (n=5) pattern having a precision of 75.2%, a recall of 74.84%, an accuracy of 74.83% and an f-measure of 67.2%. There are 67.94% labelled citation sentences that are accurately covered within bi-phrase (n=2), 71.2% for the tri-phrase (n=3) and 74.83% for the penta-phrase (n=5) after the noun phrase. This depicts that aspect-based sentiment classification calculated using forthcoming 5 words of the identified aspect obtains good results as compared to lower-order n-gram patterns. The higher-order n-grams (n=3, 5) achieved the best results than the lower order n-grams (n=2). On the contrary, the lower order n-grams obtained the worst average success rate with a precision of 68.05%, a recall of 67.98%, an accuracy of 67.94% and an f-measure of 55.18% respectively. The one possible reason for the low accuracy of the bigram feature is because bigrams are sparse and sparsity results in reducing the accuracy.

The scrutinizing of results in Table 4.9 discern that the average precision value

Table 4.9: Citation sentiment classification obtained with the 'N-Gram Before' method.

	N=2				N=3				N=5			
	P	**R**	**A**	**F**	**P**	**R**	**A**	**F**	**P**	**R**	**A**	**F**
SVM	63.2	63.2	63.24	56.2	65.9	65.3	65.34	58.4	73.7	73.3	73.32	65.9
Naïve Bayes	62.2	62.4	62.35	56.2	65.9	65.85	65.96	58.9	73.7	73.3	73.32	66.2
MaxEnt	62.2	62.34	62.35	56.2	64.45	64.12	64.89	57.1	71.6	71.6	71.63	60.15
J48	63.8	63.8	63.82	57.49	66.7	65.7	66.75	58.1	73.6	73.6	73.25	61.1
Random Forest	64.6	64.7	64.7	57.1	66.7	65.7	66.75	58.1	73.25	72.9	73.25	61.1
Average	**63.2**	**63.29**	**63.29**	**56.64**	**65.93**	**65.33**	**65.94**	**58.12**	**73.17**	**72.94**	**72.95**	**62.89**

is higher (i.e., less false positive) as soon as we move towards the higher-order n-grams which means that model has the ability to classify relevant positive citations. More concretely, the n-gram after method performs comparatively better as compared to the n-gram before method. From both n-gram after and before methods, the performance of the n-gram before parameter is worse in terms of both, higher-order and lower n-gram parameters. The worst average result is obtained with n-gram=2 having an accuracy of 63.29%, a precision of 63.2%, a recall of 63.39% and an f-measure of 57.24%. Finally, the average result of the n-gram before method for higher-order n-gram i.e., n=5 is also less than both n-gram after and around method with a precision of 73.17%, recall of 72.94%, an accuracy of 72.95% and an f-measure of 62.89%. However, the results for the higher-order n-grams show an improvement in the precision as compared to lower order n-grams both for the n-gram before and after method. This means that considering five previous words before the identified aspect produces the best result for 'n-gram before'. Similarly, in the case of 'n-gram after' method considering the next five words after the identified aspect generates the best result. Conversely, the accuracy remains almost the same for bigram and trigram patterns in the case of the before and after method. In the case of n-gram around method considering more words that precede and follow the identified aspect is important.

Specifically, the results obtained through the 'n-gram around' method are more deterministic with reference to aspect-based citation sentiment classification as compared to 'n-gram before' and 'n-gram after' parameters as shown in Table 4.10. A more concrete analysis depicts that for the citation sentiment classification of

Table 4.10: Citation sentiment classification obtained with the 'N-Gram Around' method.

	N=2				N=3				N=5			
	P	**R**	**A**	**F**	**P**	**R**	**A**	**F**	**P**	**R**	**A**	**F**
SVM	73.7	73.6	73.59	63.5	77.4	77.7	77.65	68.9	85.6	85.2	85.24	79.8
Naïve Bayes	73.1	73	72.96	62.7	74.9	74.7	74.73	64.6	82	81.5	81.52	73.7
MaxEnt	73.1	72.8	72.81	62.6	75.1	74.7	74.68	65.2	78.1	78	78.04	69.5
J48	74	74	73.95	62.9	76.6	76.6	76.59	66.4	80.7	75.1	80.71	72.1
Random Forest	75.4	75.4	75.41	64.8	77.9	77.1	77.12	70.1	83.8	84	84.03	82.8
Average	**73.86**	**73.76**	**73.74**	**63.3**	**76.38**	**76.16**	**76.15**	**67.04**	**82.04**	**80.76**	**81.91**	**75.58**

precision, recall, accuracy and f-measure it obtained an average score of 82.04%, 80.76%, 81.91% and 75.58% respectively when the n-gram parameter was set to 5. It can be observed from Figure 4.13 that higher-order n-grams outperform the lower order n-grams in all three cases i.e., forward, backward and around. Thus an optimal way to perform the citation sentiment classification is the 'n-gram around' method. The main reason for this observation is that this method contains sentiment rich phrases and also captures contextual information. This also signifies that considering both the previous and forthcoming words of the identified aspect is very important for sentiment classification. From the results presented above, we can also evaluate the performance of implied classifiers on the citation corpus. It can be seen from the results that SVM achieved the consistent performance for all the classifications using different types of n-gram based features. The maximum accuracy i.e., 85.24% is achieved by using the Support Vector Machine classifier. SVM performed better because of the clear separation margin between polarity class labels and because of its memory efficiency. One of the benefits of using SVM is it performs equally well both with structured and unstructured dataset. However, it is difficult to interpret and understand the variable weights and its impact along with final model. For the classifiers belonging to the decision tree group the performance of n-gram approach based on the Random Forest with an accuracy of 84.03% is better compared with the approach based on the J48 with an accuracy of 80.71%. Similarly, from the probabilistic classifier group, the performance of Naïve Bayes is better with an accuracy of 81.52% for n-gram around parameter setting n=5 as compared to maximum entropy with an accuracy of 78.04% because of features independence.

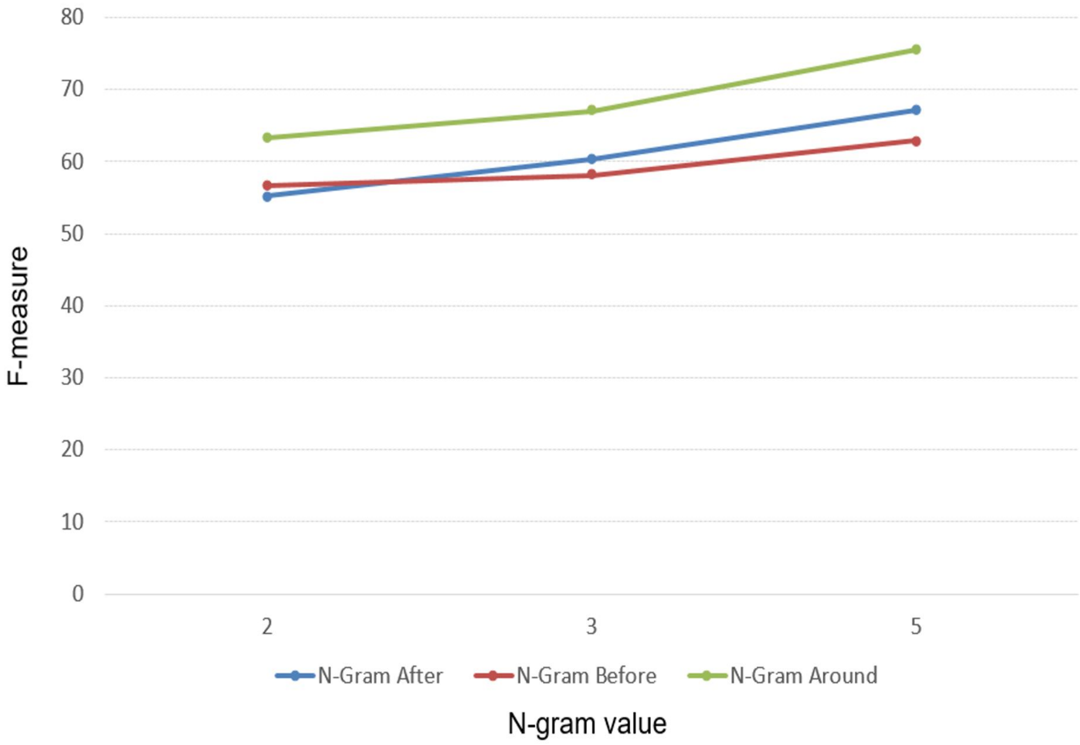

FIGURE 4.13: Aspect based citation sentiment analysis using N-gram methods.

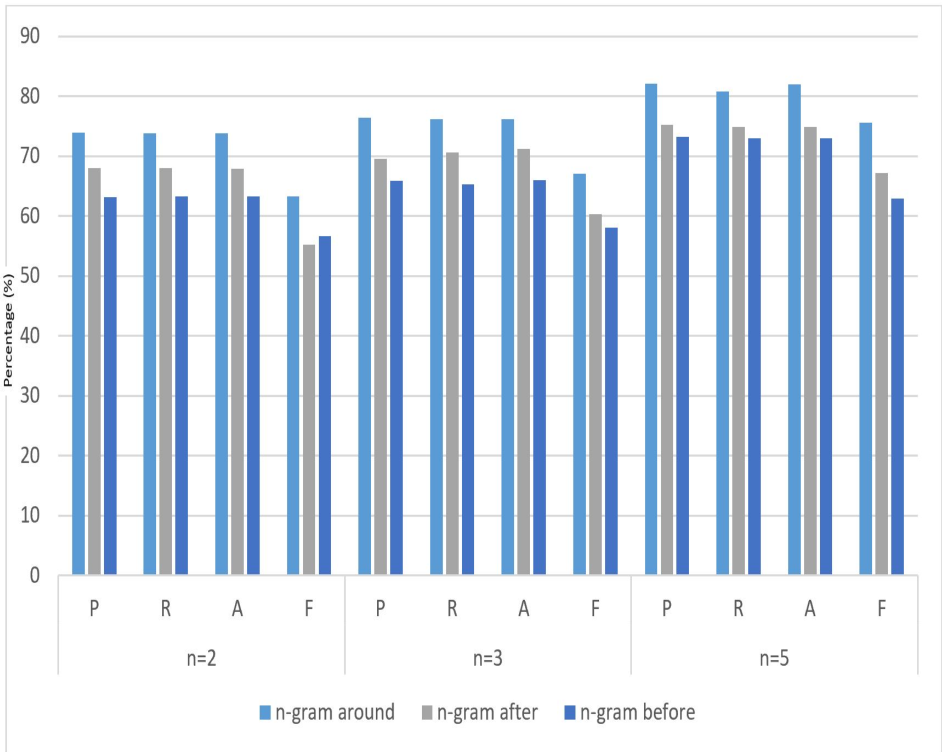

FIGURE 4.14: Citation Sentiment Classification

4.3 Context aware citation sentiment distribution

4.3.1 Citation sentiment distribution for the IMRaD structure (RQ4)

I have presented the results according to different sections of IMRaD structure in which citation contexts are most frequently distributed which is RQ4 of our research study related to the second research problem described in Chapter 1 of the thesis. Citation contexts in different sections of the research paper tend to serve different purposes. Citations in the "Introduction" section of the paper establish the context or set the stage for the intended study, establish its background and relatedness to the previous research. Similarly, citation contexts in the "Methodology" section of the paper instigate a justification for the method to be used or modified. Citation contexts in the "Results" section usually discuss the comparison of the findings with reference to the earlier studies. Citation mentions in the "Discussion" section would highlight the significance and limitations associated with the findings. As depicted in the figure, from all the 2109 citation contexts 780 (36.98%) of the citation contexts are located in the "Introduction" section, 278 (13.18%) of these citations are distributed in the "Method" section or in Section II; 696 (33%) are located in the "Literature" section, 90 (4.2%) citation contexts are in the sections of "Result"; and 316 (14.98%) in "Discussion". It can be observed that a relatively larger number of citation contexts i.e., 37 percent exist in the introduction section in this corpus. This pattern of section distribution of citation contexts adheres to our prospect on citation location distribution since the authors are most likely to cite more in the "Introduction" section and it is a widely accepted fact. The "Introduction" section of the research paper usually starts by referencing the established research, followed by referencing to most related and recent research at the end of section I. However, the "Method" section of the paper usually contains the older references, whereas the latest state-of-the-art literature is usually distributed throughout the rest of the research paper for comparative

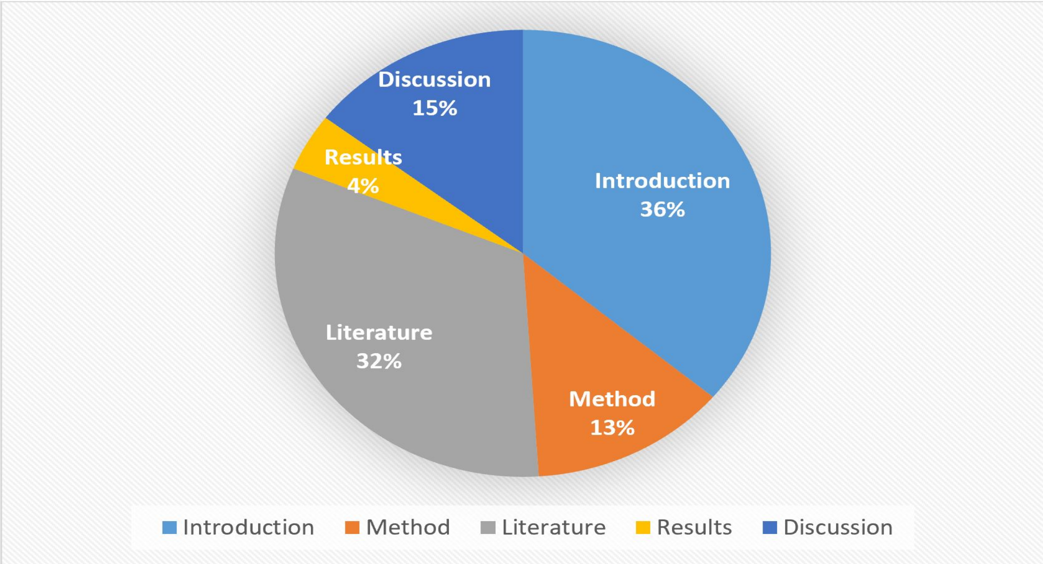

FIGURE 4.15: Descriptive statistics for citation distribution across IMRaD.

analysis. Keeping in view the functions, different article sections intend to serve, I hypothesize that the location of citation context and the sentiment associated with these citation contexts in different sections have a different impactful meaning. Therefore, I highlighted the behavior of the distribution of the sentiment across these sections and studied the intention of citing authors sentiment across different rhetorical sections.

I explored the distribution of citations sentiment in terms of IMRaD structure and represented it as a function of text progression. Fig. 4.16 represents the distribution of citation sentiment along with the article structure. The horizontal axis presents the polarity distribution of the citation sentences following the IMRaD structure. If we consider the sentiment polarity distribution, <positive>, <negative>and <objective>over the rhetorical structure, it can be observed that there is an expression of the agreement between the citing authors. As depicted in the figure, the polarity class <positive>is most frequent in the "Introduction" section of the paper and in general its frequency tend to diminish along with the "Methods", "Literature", "Results" and "Discussion" sections. This shows that the expression of positive sentiment towards the cited paper is most common at the start of the research paper especially in the "Introduction" section followed by the "Results" and "Discussion" section. The same trend is observed by the

<negative>polarity class having the highest frequency in the “Introduction” and “Discussion” section along with a steady fall along with the “Methods”, “Literature” and “Results” sections. However, it can also be observed that the frequency of negative citations in the discussion section supersedes the presence of positive and neutral citations. The descriptive statistics regarding the positive and negative citation sentiment is reported in the table and article sections are ranked as per the frequency in the descending order. An important rhetorical function of citation context is demonstrated by the strong presence of <positive>and <negative>class in the “Introduction” and “Discussion” sections of the paper. This tendency gives an idea as to where most sentiment bearing citation contexts could be found and papers cited in these sections should be considered having different sentiment strengths ascertaining to our hypothesis, not all the citation sentiments are equal. A general trend of concentration of more sentiment bearing citation mention is depicted at the commencement of the document with a further rapid decrease to 25%, where it essentially smooths out, trailed by another peak for the objective citations in the “Literature” section before diminishing at the end the research paper. Additionally, for the <objective>class the results are consistent with the expectation, objective citation contexts are more concerted in the second section which is “Literature” in a vast majority of cases. The comparison of the curves shows the distribution of citation contexts is similar across all the classes through the “Method” section is characterized with having a relatively smaller proportion of sentimental citation contexts which relishes bigger in the “Results” section and gets higher in the “Discussion” section. However, the most significant result is that the “Discussion” section is designated with the largest number of negative citation contexts as compared to “Results” and “Introduction” sections which associates a new property to it not yet described in the scientific literature. The findings regarding the “Method” section are consistent with that of (Bertin et al. 2016; Hashimoto et al. 2016) which shows the atypical nature of this section as compared to other rhetorical sections of the research paper. Our cognitive hypothesis was the distribution of the citation sentiment in scientific articles is affected by the structure of the articles represented through IMRaD sections. We

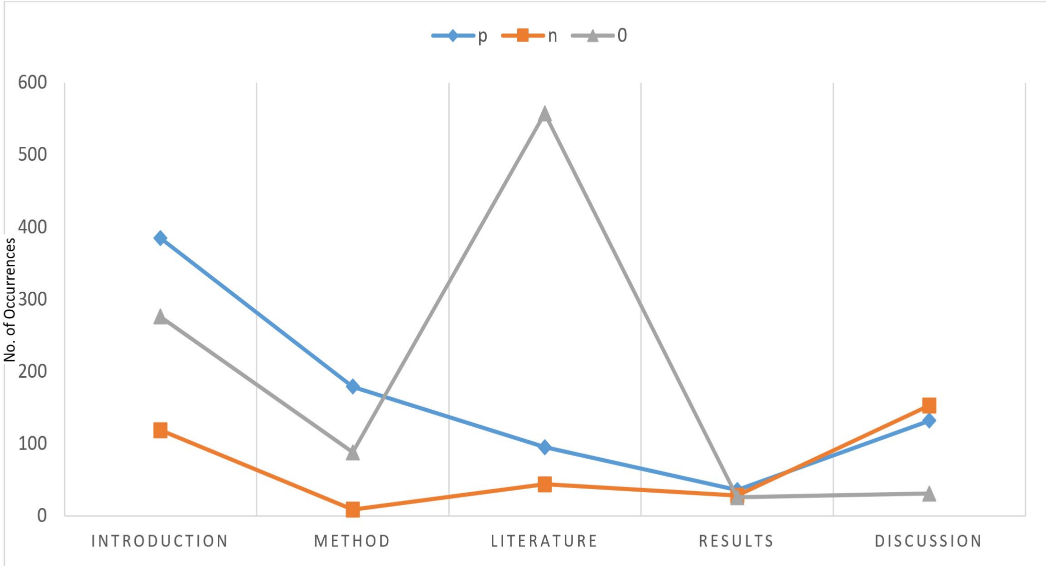

FIGURE 4.16: Distribution of citation sentiment along the IMRaD structure.

considered the relationship between the argumentative purpose of each rhetorical section and the distribution of citation sentiment. The results demonstrated a strong relationship between the argumentative structure of the research paper and sentiment expression across them. The observed distribution of citation sentiment across different sections of the paper asserts the results reported in previous studies which have shown that more specifically the first sections contains the highest frequency of sentiment bearing citation contexts. I also divulge the uncertainty with reference to the polarity of the citation context and location of the citation mention which is reported in Table 4.11. The results reveal that the highest level of uncertainty is for the "Discussion" section (mean=.25), the second-highest is for the "Introduction" section (mean= .19), the lowest value of uncertainty is for the "Method" section (mean=0.05) and the second-lowest value is for the "Literature" section (mean=0). Our results provide clear evidence the distribution of sentiment across sentiment bearing citation sentences and objective sentences is not homogeneous. The distribution of <positive>and <negative>citation sentences monotonously diminishes in the "Literature" section of the papers whereas, high density peaks are more frequent in the "Literature" section for the <objective>citation sentences. This trend is a bit different in regards to the other sections of the research article for predictability of the citation sentiment.

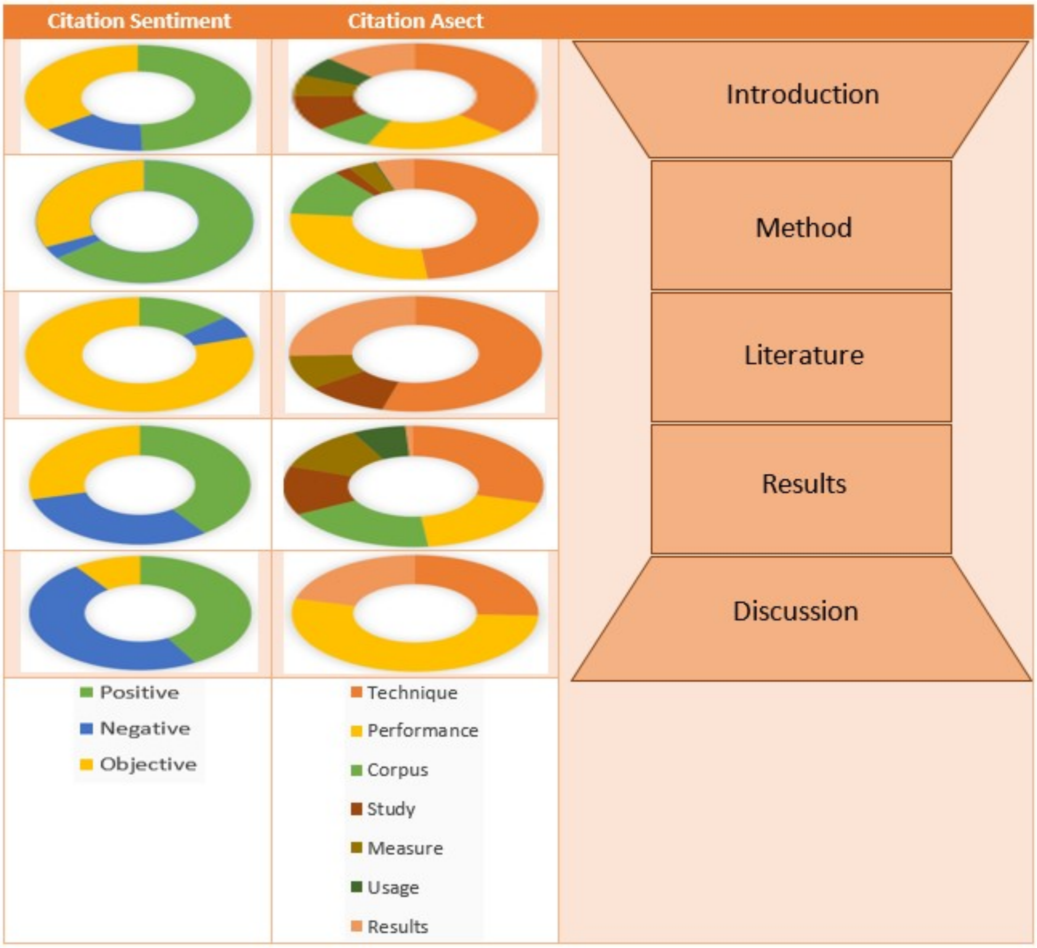

FIGURE 4.17: Detailed view of citation aspect and sentiment distribution along the IMRaD structure.

TABLE 4.11: Descriptive Statistics of Citation Sentiment Uncertainty across IMRaD.

Section	N	Minimum	Maximum	Mean	Std. Deviation
Introduction	780	-0.1	1.52	.10	.49
Method	302	-0.1	.78	.01	.15
Literature	696	-0.1	1.42	.15	.42
Results	90	-0.1	1.09	.19	.38
Discussion	216	-0.1	1.46	.25	.40

TABLE 4.12: Descriptive Statistics for Citation Sentiment (Positive) across IMRaD.

Section	N	Minimum	Maximum	Mean	Std. Deviation
Introduction	385	0	3	.67	.75
Method	179	0	5	.45	.65
Literature	132	0	6	.40	.82
Results	95	0	3	.15	.40
Discussion	36	0	1	.07	.19

TABLE 4.13: Descriptive Statistics for Citation Sentiment (Negative) across IMRaD.

Section	N	Minimum	Maximum	Mean	Std. Deviation
Introduction	153	0	3	.11	.39
Method	119	0	5	.09	.48
Literature	44	0	4	.03	.15
Results	28	0	2	.02	.12
Discussion	9	0	1	.01	.00

4.3.2 Impact of citation contextual window on determining the citation sentiment (RQ6)

After having studied the distribution of citation sentiment across the rhetorical structure, the next step is determining how sentiment varies over the citation context and how diversified is the sentiment contained in the contextual sentences (RQ6). For this sake, we have to return to the citation sentences individually to understand the significance of the context in which they appeared. So, in this section, the results are provided regarding the performance of different length contextual citation sentences and the impact of their polarity on the polarity of citation sentences. The polarity of citation sentence is compared as opposed to the polarity of 3sent i.e., the annotated citation sentence plus one sentence before and after the citation sentence (non-citation sentence) and 5sent i.e., the citation sentence plus two sentences before and after the citation sentence (non-citation sentence). To investigate whether neighboring or surrounding sentences might be an impactful indicator for identifying the polarity of objective citation sentences these methods were devised. The pattern of contextual sentences in the vicinity of the citation sentence is examined for highlighting the sentiment variability in the framing sentences using the citation context which can better help in producing the articles summaries. Further, I have made a comparison between the polarity of positive, negative and objective citations and the impact of contextual polarity on citation polarity separately. These patterns confirm our hypothesis that merely using the citation sentence would not be enough in identifying citation sentiment because authors usually hide the sentiment as duplicitous praise. I calculated the number of positive, negative and neutral instances divided by the total sample

TABLE 4.14: Positive Citations followed by Different Sentiment Context.

Citation Text	Polarity	POST Text	Polarity
"[] has recently proposed a simpler SVM-based algorithm for analogical classification called Pair-Class."	P	"While it does not adopt a set-based or distributional model of relational similarity."	N
"The SS model with e2e info is very similar to selfish routing."	P	"Where each flow tries to minimize its average latency over multiple paths without coordinating with other flows."	N
"We use evaluations similar to those used before []."	P	"However, whereas most existing studies use only one dataset, or hand selected parts thereof."	N
"[] conducted the first study of general non-atomic congestion games and showed a tight bound of 4/3 for the price of anarchy with linear latency functions."	P	"In the absence of prices the decentralized equilibrium can be highly inefficient."	N

along with each window size and the average sentiment score of the contextual sentences following and proceeding the citation sentence to understand the uniformity or variability. The emerged patterns ascertain a sentiment drift in the context of positively, negatively and objectively cited research papers. One of the key findings from the identified patterns is that for the majority of positive class citations the contextual text surrounding the citation text is negative. This indicates that for most of the cases pertaining to positive citations the post text bears a negative sentiment orientation which determines that without a wider view of citation context it is difficult to comprehend if a citing sentence is, in fact, positive or negative. Some of the examples for this scenario are previewed in Table 4.14.

Whereas, for the negative class citations, there is assimilation in the sense that the text following the citation text is negative. There is a pattern of opinion agreement between the citation sentence and the contextual framing sentences which can be previewed from table 4.15. Neighboring sentiments tend to agree on the citation sentiment for a majority of the cases belonging to the negative sentiment class i.e., 60% of such claims shared the same polarity. This means that if an author cites a paper explicitly in a negative way then the forthcoming text is of the same opinion orientation and there is no need to hide the citation with the

TABLE 4.15: Negative Citations followed by Negative Sentiment Context.

Citation Text	Polarity	POST Text	Polarity
"Automatic text summarization approaches have offered reasonably well-performing approximations for identifying important sentences [] but, not surprisingly, text (re)generation has been a major challenge despite some work on sub-sentential modification []"	N	"An additional drawback of extractive approaches is that estimates for the importance of larger text units such as sentences depend on the length of the sentence []."	N
"The POS disambiguation has usually been performed by statistical approaches mainly using hidden markov model (HMM) []."	N	"However, since statistical approaches take into account neighboring tags only within a limited window (usually two or three), sometimes the decision cannot cover all linguistic contexts necessary for POS disambiguation."	N
"[] recently advocated the need for a uniform approach to corpus-based semantic tasks."	N	"[] recasts a number of semantic challenges in terms of relational or analogical similarity."	N

TABLE 4.16: Objective Citations followed by Different Sentiment Context .

Citation Text	Polarity	POST Text	Polarity
"[] has proposed an estimation method for the N-gram language model using the Baum-Welch re-estimation algorithm [] from an untagged corpus and []."	O	"They have applied this method to an English tagging system but resulted in less accuracy."	N
"The two systems we use are ENGCG [] and the Xerox Tagger []."	O	"We discuss problems caused by the fact that these taggers use different tag sets, and present the results obtained by applying the combined taggers to a previously unseen sample of text."	N
"POS disambiguation has usually been performed by statistical approaches, mainly using the hidden Markov model (HMM) in English research communities []."	N	"And the approaches are also dominant for Korean with slight improvements for the agglutinative nature of Korean."	P

duplicitous praise. However, the situation is different in the case of the positive and neutral citations where the authors avoid to contradict or criticize the cited work explicitly and tend to write the criticism in the forthcoming sentences and not in the citation sentence.

Sentiment convergence would imply that as the number of contextual citation sentences for a cited research paper grow, most neighboring sentences for positive citations diverge. A general trend between the length of contextual sentences and their impact on identifying the polarity of citation sentence has been found. It can be seen from Table 4.16 that the proportion of sentiment drift between the citation sentence polarity and the contextual sentence polarity decreases as more contextual sentences are considered as a combined case of positive, negative and objective citation sentences. Our explanation for this particular trend led us to the conclusion that up to a point of 3sent contextual window, the aspects or sentiment for what the paper is cited exists but beyond that point, the contextual sentences represent the same few things. The analysis of the results further reveal that sentence-based citation contexts tend to be more effective in determining the citation polarity of objective citations. However, for the positive and objective citations considering more contextual sentences would be more helpful for the citation polarity identification: 1) for determining the polarity of objective citations as compared to positive or negative citations; 2) as in the majority of the positively cited references the sentiment is hedged.

Chapter 5

Conclusion and Future Work

5.1 Introduction

The contribution of automated citation sentiment analysis is twofold. Initially, it has explored the effectiveness of using adjectives, adverbs, and their combinations for document-level sentiment classification of the citation text using a gold standard citation sentiment corpus which is our first research problem and research question *RQ1: What is the impact of higher-order POS and word-based n-grams in detecting the citation sentiment?*. Afterwards, it has investigated the efficacy of applying the commercial tools on the citation corpus for sentiment detection. The analysis of the results has revealed that high order n-grams ($n = 5$) for adjective and adverb combinations play a major role in improving the accuracy of the sentiment classification. The outcomes of the study have shown an accuracy of above 80% against the manually annotated corpus using higher-order n-gram adjective and adverb combinations. The comparative analysis of the results with the state of the art research based on using n-grams of different lengths deduced that unigram, bigram, and trigram performed better when used in combination as compared to unigram and unigram plus bigrams. The experimental results have also determined that current sentiment analysis tools more specifically THEYSAY are not efficient enough to detect sentiments accurately as the majority of the citations comprised of multiple sentences. A possible direction for future work could be the contemplation of more features and dependency relationships for citation sentiment classification using different machine learning algorithms. The availability of unlimited sources of scientific information plethora on the web has complicated the process of finding relevant research papers in the contemporary academic setups. The currently designed systems are primarily suitable for adhering ad-hoc information needs wherein the accumulation and establishing a connection between the papers are the tasks that are left to be tackled by the research community. After studying the performance of citation sentiment classification using higher order n-gram word phrases I presented the aspect based citation sentiment classification approach which is a qualitative bibliometric pattern of research impact evaluation complementing the citation aspects with its purpose and polarity *(RQ2: How the type of n-gram model effect the aspect-based sentiment classification?).* This is the third research problem identified and to the best of our knowledge, this paper is among the first that tackles the issue of detecting and analysing aspects from the citation text to foster and decipher citation performances. Connecting technical aspect terms to the sentiments can be used for many purposes, including the creation of aspect-based sentiment profile of the paper, identifying meaningful aspects of the cited work, how sentiment differs from one aspect to another, how sentiment evolves over a collaborative-dynamic evolving paradigm. Firstly, I have proposed a qualitative aspect based citation sentiment classification scheme by using aspect–opinion phrase patterns based on linguistic rules combined with synonyms and chi-square test to effectively detect aspects concerning the cited work. According to research findings, “methodology”, “performance”, “corpus”, “study”, “measure” and “result” are the most frequently discussed aspects of a research study. This work demonstrates the feasibility of automated analysis for identifying the ‘material’ aspects or features associated with the cited work combining different extraction criterions like frequency-based, opinion-based, rule-based and aspect-ranking. Secondly, I considered different length n-grams after, before and around the noun phrase for identifying the sentiment polarity on a specific citation aspect and evaluated the citation sentiment classification accuracy *(RQ5:*

How varied sized contextual window effect citation sentiment detection?). The experimental results revealed that the best average results are obtained by the 'n-gram around' method. This also ascertains that higher-order n-grams facilitate in decimating the citation aspect polarity in a mixture context. Such an observation has never been testified on a large-scale dataset before. The citation sentences classification is performed using different machine learning algorithms i.e., Support Vector Machine (SVM), Naïve Bayes (NB), Maximum Entropy, J48 and Random Forest *(RQ3: How can different machine learning techniques contribute to aspect-based sentiment detection?).* The proposed approach has obtained encouraging results for aspect sentiment classification with an accuracy of 0.819, a precision of 0.82, a recall of 0.807 and an f-measure of 0.755. SVM achieved consistent performance for all the classifications using different types of n-gram based features. The maximum accuracy i.e., 85.24% is achieved by using the Support Vector Machine classifier. The analysis of the results depicts that implying a discriminating classifier along with higher-order n-grams can achieve comparable performance. From the decision tree classifiers group, Random Forest performed better with an accuracy of 84.03% and from the probabilistic classifier group, Naïve Bayes performed better with an accuracy of 81.52%. In future, I intend to examine the application of a hybrid feature selection method to further reduce the number of noisy n-grams. In regards to context-based citation analysis which is the second research problem of the study, I have proposed that the citation sentiment can be determined by analyzing the rhetorical intent of the citation context that surrounds the citation text *(RQ4: What is the pattern of citation sentiment distribution across rhetorical structure of research paper?).* The investigations made by our study have determined that citation contexts, rhetorical structure and citation sentiment are related in important ways. The analysis of the results depicts that the "Discussion" section encompasses the largest number of negative citation contexts as compared to other sections of the research paper. For *RQ6: What is the effect of citation context on citation polarity and purpose classification?* I have also shown that the context in which the citation exists has a significant effect on the polarity of the citation sentence. In particular, comparison regarding the range of citation contexts depict a general trend that longer the citation context, i.e., considering more sentences, more efficiently citation sentiment is identified which means the better the performance. The results have also shown the optimal citation context which seems to be different for the polarized citation sentences and neutral sentences. For the positive or negative citation sentences considering 3sent would be enough however for the objective citations considering more contextual sentences like 5sent would be helpful. However, as a future work, it would be interesting to study the longer sentence based and window based citation contexts by considering more linguistic clues and contextual terms.

5.2 Future Work

The research work presented in this dissertation can be further extended in the following possible directions:

1. Since the corpuses used in the research study are imbalanced and small-sized, in future I intend to use some state-of-the-art methods for imbalanced learning and improving the performance of minority (negative or neutral) classes which can affect the classifier's performance. Another aspect is selecting classifiers, for some of the classification tasks I intend to use novel deep learning and topic modeling algorithms and compare the findings.

2. A potential future research direction for aspect-based citation analysis is mapping implicit citation aspects to explicit aspects and author-specific aspect sentiments. For this, I will be leveraging the use of already identified opinion chunks from the citation sentences using the ConceptNet and Similarity Index.

3. For citation context analysis, our future work will examine the correlations between the nature of the references in terms of the published year, subject category, published venue, publication source, article type etc., semantic features and position of the citation text in the cited research paper. This will

also encompass mining other features for studying the relationship between citation sentence, cited text span and its location. I also plan to accomplish an additional analysis on citation contexts and citations across varied scientific domains and cross-field comparisons to generalize the findings of this research study and report differences in citing behavior. This will be coupled with performing a weighted citation analysis by assigning weights to each citation mention based on their appearance in the cited research paper section, mention frequency and citation count.

5.3 Utility of Citation Sentiment Analysis

This section describes a number of ways in which Citation Sentiment Analysis can be helpful for researchers. One of the potential applications is assigning different weights to existing count-based bibliometrics measures leveraging citation sentiment. This will result in determining an unbiased citation count and help in mitigating better estimation of the impact of cited paper. New ranking measures can be devised by combining sentiment, frequency and link analysis of citations for creating a more efficient and robust qualitative measure. Another application of the proposed work in identifying the hedging and identifying personal bias by analysing the trends of appraisal and criticism. Further, we can use positive citations in identifying the research contributions in the form of innovations and improvements that the cited paper has made in a specific domain. Whereas negative citations can be used for unveiling the potential gaps and unaddressed issues in the existing research approaches. All of these utilities of citation sentiment analysis are real problems in the domain of citation analysis. Developing applications by amalgamating the citation sentiment analysis with research paper indexing measures will be beneficial for research community.

Bibliography

[1] A. Athar, "Sentiment analysis of citations using sentence structure-based features," in *Proceedings of the ACL 2011 student session.* Association for Computational Linguistics, 2011, pp. 81–87.

[2] G. Parthasarathy and D. Tomar, "Sentiment analyzer: Analysis of journal citations from citation databases," in *2014 5th international conference-confluence the next generation information technology summit (confluence).* IEEE, 2014, pp. 923–928.

[3] S. Stamou, N. Mpouloumpasis, and L. Kozanidis, "Deriving the impact of scientifi c publications by mining citation opinion terms." *JDIM*, vol. 7, no. 5, pp. 283–289, 2009.

[4] M. Hernández and J. M. Gómez, "Sentiment, polarity and function analysis in bibliometrics: a review," *Natural Language Processing and Cognitive Science*, p. 149, 2015.

[5] I. C. Kim and G. R. Thoma, "Automated classification of author's sentiments in citation using machine learning techniques: A preliminary study," in *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB).* IEEE, 2015, pp. 1–7.

[6] J. Oh, H. Chang, J. A. Kim, M. Choi, Z. Park, Y. Cho, and E.-g. Lee, "Citation analysis for biomedical and health sciences journals published in korea," *Healthcare informatics research*, vol. 23, no. 3, pp. 218–225, 2017.

[7] N. Tandon and A. Jain, "Citation context sentiment analysis for structured summarization of research papers," in *35th German conference on artificial intelligence*. Citeseer, 2012, pp. 24–27.

[8] R. Klavans and K. W. Boyack, "Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?" *Journal of the Association for Information Science and Technology*, vol. 68, no. 4, pp. 984–998, 2017.

[9] A. Athar and S. Teufel, "Context-enhanced citation sentiment detection," in *Proceedings of the 2012 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies*. Association for Computational Linguistics, 2012, pp. 597–601.

[10] H. Gurzki and D. M. Woisetschläger, "Mapping the luxury research landscape: A bibliometric citation analysis," *Journal of Business Research*, vol. 77, pp. 147–166, 2017.

[11] H. Liu, "Sentiment analysis of citations using word2vec," *arXiv preprint arXiv:1704.00177*, 2017.

[12] Z. Ma, J. Nam, and K. Weihe, "Improve sentiment analysis of citations with author modelling," in *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2016, pp. 122–127.

[13] B. H. Butt, M. Rafi, A. Jamal, R. S. U. Rehman, S. M. Z. Alam, and M. B. Alam, "Classification of research citations (crc)," *arXiv preprint arXiv:1506.08966*, 2015.

[14] P. Kazi, M. Patwardhan, and P. Joglekar, "Towards a new perspective on context based citation index of research articles," *Scientometrics*, vol. 107, no. 1, pp. 103–121, 2016.

[15] C. Dong and U. Schäfer, "Ensemble-style self-training on citation classification," in *Proceedings of 5th international joint conference on natural language processing*, 2011, pp. 623–631.

[16] C. Jochim and H. Schütze, "Towards a generic and flexible citation classifier based on a faceted classification scheme," in *Proceedings of COLING 2012*, 2012, pp. 1343–1358.

[17] X. Zhu, P. Turney, D. Lemire, and A. Vellino, "Measuring academic influence: Not all citations are equal," *Journal of the Association for Information Science and Technology*, vol. 66, no. 2, pp. 408–427, 2015.

[18] X. Wan and F. Liu, "Are all literature citations equally important? automatic citation strength estimation and its applications," *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1929–1938, 2014.

[19] H. Xu, E. Martin, and A. Mahidadia, "Using heterogeneous features for scientific citation classification," in *Proceedings of the 13th conference of the Pacific Association for Computational Linguistics*, 2013.

[20] S. D. Tembhurnikar and N. N. Patil, "Topic detection using bngram method and sentiment analysis on twitter dataset," in *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)*. IEEE, 2015, pp. 1–6.

[21] H. Alam, A. Kumar, T. Werner, and M. Vyas, "Are cited references meaningful? measuring semantic relatedness in citation analysis." *BIRNDL@ SIGIR (1)*, vol. 1888, pp. 113–118, 2017.

[22] B. Yu, "Automated citation sentiment analysis: What can we learn from biomedical researchers," *Proceedings of the American Society for Information Science and Technology*, vol. 50, no. 1, pp. 1–9, 2013.

[23] J. Xu, Y. Zhang, Y. Wu, J. Wang, X. Dong, and H. Xu, "Citation sentiment analysis in clinical trial papers," in *AMIA annual symposium proceedings*, vol. 2015. American Medical Informatics Association, 2015, p. 1334.

[24] A. Athar, "Sentiment analysis of scientific citations," University of Cambridge, Computer Laboratory, Tech. Rep., 2014.

[25] S.-U. Hassan, A. Akram, and P. Haddawy, "Identifying important citations using contextual information from full text," in *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 2017, pp. 1–8.

[26] M. Bertin and I. Atanassova, "Weak links and strong meaning: The complex phenomenon of negational citations." in *BIR@ ECIR*, 2016, pp. 14–25.

[27] K. Hashimoto, T. Soonklang, and S. Hirokawa, "Feature words of moves in scientific abstracts," in *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, 2016, pp. 144–149.

[28] H. Saggion, A. AbuRa'ed, and F. Ronzano, "Trainable citation-enhanced summarization of scientific articles," in *Cabanac G, Chandrasekaran MK, Frommholz I, Jaidka K, Kan M, Mayr P, Wolfram D, editors. Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL); 2016 June 23; Newark, United States.[place unknown]: CEUR Workshop Proceedings; 2016. p. 175-86.* CEUR Workshop Proceedings, 2016.

[29] J. Bar-Ilan and G. Halevi, "Post retraction citations in context: A case study," *Scientometrics*, vol. 113, no. 1, pp. 547–565, 2017.

[30] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels' reviews," *Journal of computational science*, vol. 27, pp. 386–393, 2018.

[31] G. C. Nobre and E. Tavares, "Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study," *Scientometrics*, vol. 111, no. 1, pp. 463–492, 2017.

[32] X. Liu, Y. Yu, C. Guo, Y. Sun, and L. Gao, "Full-text based context-rich heterogeneous network mining approach for citation recommendation," in *IEEE/ACM joint conference on digital libraries.* IEEE, 2014, pp. 361–370.

[33] R. Meng, W. Lu, Y. Chi, and S. Han, "Automatic classification of citation function by new linguistic features," *iConference 2017 Proceedings*, 2017.

[34] D. Anand and D. Naorem, "Semi-supervised aspect based sentiment analysis for movies using review filtering," *Procedia Computer Science*, vol. 84, pp. 86–93, 2016.

[35] M. Hernández Álvarez, J. M. Gómez, P. Martínez-Barco *et al.*, "Annotated corpus for citation context analysis," 2016.

[36] A. M. Abirami and A. Askarunisa, "Feature based sentiment analysis for service reviews." *J. UCS*, vol. 22, no. 5, pp. 650–670, 2016.

[37] X. Zhang, H. Chen, W. Wang, and P. Ordóñez de Pablos, "What is the role of it in innovation? a bibliometric analysis of research development in it innovation," *Behaviour & Information Technology*, vol. 35, no. 12, pp. 1130–1143, 2016.

[38] S. Ghosh, D. Das, and T. Chakraborty, "Determining sentiment in citation text and analyzing its impact on the proposed ranking index," in *International Conference on Intelligent Text Processing and Computational Linguistics.* Springer, 2016, pp. 292–306.

[39] H. Kilicoglu, "Biomedical text mining for research rigor and integrity: tasks, challenges, directions," *Briefings in bioinformatics*, vol. 19, no. 6, pp. 1400–1414, 2018.

[40] M. Valenzuela, V. Ha, and O. Etzioni, "Identifying meaningful citations," in *Workshops at the twenty-ninth AAAI conference on artificial intelligence*, 2015.

[41] A. Khalid, F. Alam, and I. Ahmed, "Extracting reference text from citation contexts," *Cluster Computing*, vol. 21, no. 1, pp. 605–622, 2018.

[42] C. Lu, Y. Ding, and C. Zhang, "Understanding the impact change of a highly cited article: a content-based citation analysis," *Scientometrics*, vol. 112, no. 2, pp. 927–945, 2017.

[43] A. Yousif, Z. Niu, J. K. Tarus, and A. Ahmad, "A survey on sentiment analysis of scientific citations," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1805–1838, 2019.

[44] I. Tahamtan and L. Bornmann, "What do citation counts measure? an updated review of studies on citations in scientific documents published between 2006 and 2018," *Scientometrics*, vol. 121, no. 3, pp. 1635–1684, 2019.

[45] A. Cohan, W. Ammar, M. van Zuylen, and F. Cady, "Structural scaffolds for citation intent classification in scientific publications," *arXiv preprint arXiv:1904.01608*, 2019.

[46] Y. Wang, A. J. Bowers, and D. J. Fikis, "Automated text data mining analysis of five decades of educational leadership research literature: Probabilistic topic modeling of eaq articles from 1965 to 2014," *Educational administration quarterly*, vol. 53, no. 2, pp. 289–323, 2017.

[47] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: The next generation of citation analysis," *Journal of the association for information science and technology*, vol. 65, no. 9, pp. 1820–1833, 2014.

[48] D. Zhao, A. Cappello, and L. Johnston, "Functions of uni-and multi-citations: Implications for weighted citation analysis," *Journal of Data and Information Science*, vol. 2, no. 1, pp. 51–69, 2017.

[49] A. Khalid, F. A. Khan, M. Imran, M. Alharbi, M. Khan, A. Ahmad, and G. Jeon, "Reference terms identification of cited articles as topics from citation contexts," *Computers & Electrical Engineering*, vol. 74, pp. 569–580, 2019.

[50] Z. Taşkın and U. Al, "A content-based citation analysis study based on text categorization," *Scientometrics*, vol. 114, no. 1, pp. 335–357, 2018.

[51] H. Kilicoglu, Z. Peng, S. Tafreshi, T. Tran, G. Rosemblat, and J. Schneider, "Confirm or refute?: A comparative study on citation sentiment classification in clinical research publications," *Journal of biomedical informatics*, vol. 91, p. 103123, 2019.

[52] A. Yousif, Z. Niu, J. Chambua, and Z. Y. Khan, "Multi-task learning model based on recurrent convolutional neural networks for citation sentiment and purpose classification," *Neurocomputing*, vol. 335, pp. 195–205, 2019.

[53] B. Fetahu, K. Markert, and A. Anand, "Fine grained citation span for references in wikipedia," *arXiv preprint arXiv:1707.07278*, 2017.

[54] D. Duma, C. Sutton, and E. Klein, "Context matters: Towards extracting a citation's context using linguistic features," in *2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL).* IEEE, 2016, pp. 201–202.

[55] S. Liu, C. Chen, K. Ding, B. Wang, K. Xu, and Y. Lin, "Literature retrieval based on citation context," *Scientometrics*, vol. 101, no. 2, pp. 1293–1307, 2014.

[56] T. Kuhn, M. Perc, and D. Helbing, "Inheritance patterns in citation networks reveal scientific memes," *Physical Review X*, vol. 4, no. 4, p. 041036, 2014.

[57] I. Fister Jr, I. Fister, and M. Perc, "Toward the discovery of citation cartels in citation networks," *Frontiers in Physics*, vol. 4, p. 49, 2016.

[58] D. Kaplan, T. Tokunaga, and S. Teufel, "Citation block determination using textual coherence," *Journal of Information Processing*, vol. 24, no. 3, pp. 540–553, 2016.

[59] A. Cohan and N. Goharian, "Scientific article summarization using citation-context and article's discourse structure," *arXiv preprint arXiv:1704.06619*, 2017.

[60] P. Knoth, P. Gooch, and K. Jack, "What others say about this work? scalable extraction of citation contexts from research papers," in *International Conference on Theory and Practice of Digital Libraries*. Springer, 2017, pp. 287–299.

[61] A. Abu-Jbara and D. Radev, "Reference scope identification in citing sentences," in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2012, pp. 80–90.

[62] J. An, N. Kim, M.-Y. Kan, M. K. Chandrasekaran, and M. Song, "Exploring characteristics of highly cited authors according to citation location and content," *Journal of the Association for Information Science and Technology*, vol. 68, no. 8, pp. 1975–1988, 2017.

[63] J. O'Connor, "Citing statements: Computer recognition and use to improve retrieval," *Information Processing & Management*, vol. 18, no. 3, pp. 125–131, 1982.

[64] A. Ritchie, S. Robertson, and S. Teufel, "Comparing citation contexts for information retrieval," in *Proceedings of the 17th ACM conference on Information and knowledge management*, 2008, pp. 213–222.

[65] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 421–430.

[66] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. L. Giles, "A neural probabilistic model for context based citation recommendation," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[67] V. Qazvinian and D. R. Radev, "Identifying non-explicit citing sentences for citation-based summarization," in *Proceedings of the 48th annual meeting of the association for computational linguistics*. Association for Computational Linguistics, 2010, pp. 555–564.

[68] Q. Mei and C. Zhai, "Generating impact-based summaries for scientific literature," in *Proceedings of ACL-08: HLT*, 2008, pp. 816–824.

[69] V. Vyas, K. Ravi, V. Ravi, V. Uma, S. Setlur, and V. Govindaraju, "Article citation study: Context enhanced citation sentiment detection," *arXiv preprint arXiv:2005.04534*, 2020.

[70] D. Mercier, S. T. R. Rizvi, V. Rajashekar, A. Dengel, and S. Ahmed, "Impactcite: An xlnet-based method for citation impact analysis," *arXiv preprint arXiv:2005.06611*, 2020.

[71] S. Nazir, M. Asif, S. Ahmad, F. Bukhari, M. T. Afzal, and H. Aljuaid, "Important citation identification by exploiting content and section-wise in-text citation count," *PloS one*, vol. 15, no. 3, p. e0228885, 2020.

[72] H. Aljuaid, R. Iftikhar, S. Ahmad, M. Asif, and M. T. Afzal, "Important citation identification using sentiment analysis of in-text citations," *Telematics and Informatics*, vol. 56, p. 101492, 2021.

[73] T. A. Rana and Y.-N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artificial Intelligence Review*, vol. 46, no. 4, pp. 459–483, 2016.

[74] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using gini index feature selection method and svm classifier," *World wide web*, vol. 20, no. 2, pp. 135–154, 2017.

[75] M. T. Ikram, N. A. Butt, and M. T. Afzal, "Open source software adoption evaluation through feature level sentiment analysis using twitter data," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, no. 5, pp. 4481–4496, 2016.

[76] I. Penalver-Martinez, F. Garcia-Sanchez, R. Valencia-Garcia, M. A. Rodriguez-Garcia, V. Moreno, A. Fraga, and J. L. Sanchez-Cervantes,

"Feature-based opinion mining through ontologies," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5995–6008, 2014.

[77] M. J. Er, F. Liu, N. Wang, Y. Zhang, and M. Pratama, "User-level twitter sentiment analysis with a hybrid approach," in *International symposium on neural networks*. Springer, 2016, pp. 426–433.

[78] O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to the sentiment analysis problem at the sentence level," *Knowledge-Based Systems*, vol. 108, pp. 110–124, 2016.

[79] D. Georgiou, A. MacFarlane, and T. Russell-Rose, "Extracting sentiment from healthcare survey data: An evaluation of sentiment analysis tools," in *2015 Science and Information Conference (SAI)*. IEEE, 2015, pp. 352–361.

[80] N. Zainuddin, A. Selamat, and R. Ibrahim, "Twitter feature selection and classification using support vector machine for aspect-based sentiment analysis," in *International conference on industrial, engineering and other applications of applied intelligent systems*. Springer, 2016, pp. 269–279.

[81] M. d. P. Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. A. Rodriguez-Garcia, and R. Valencia-Garcia, "Sentiment analysis on tweets about diabetes: an aspect-level approach," *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[82] M. d. P. Salas-Zárate, R. Valencia-García, A. Ruiz-Martínez, and R. Colomo-Palacios, "Feature-based opinion mining in financial news: an ontology-driven approach," *Journal of Information Science*, vol. 43, no. 4, pp. 458–479, 2017.

[83] H. Luna-Aveiga, J. Medina-Moreira, K. Lagos-Ortiz, O. Apolinario, M. A. Paredes-Valverde, M. del Pilar Salas-Zárate, and R. Valencia-García, "Sentiment polarity detection in social networks: an approach for asthma disease management," in *International conference on computer science, applied mathematics and applications*. Springer, 2017, pp. 141–152.

[84] B. Agarwal and N. Mittal, "Prominent feature extraction for review analysis: an empirical study," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 28, no. 3, pp. 485–498, 2016.

[85] C. Quan and F. Ren, "Feature-level sentiment analysis by using comparative domain corpora," *Enterprise Information Systems*, vol. 10, no. 5, pp. 505–522, 2016.

[86] N. Zainuddin, A. Selamat, and R. Ibrahim, "Improving twitter aspect-based sentiment analysis using hybrid approach," in *Asian conference on intelligent information and database systems.* Springer, 2016, pp. 151–160.

[87] M. Wang, D. Leng, J. Ren, Y. Zeng, and G. Chen, "Sentiment classification based on linguistic patterns in citation context," *Current Science*, vol. 117, no. 4, pp. 606–616, 2019.

[88] S. Chakraborty, P. Goyal, and A. Mukherjee, "Aspect-based sentiment analysis of scientific reviews," in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 2020, pp. 207–216.

[89] Z. Kastrati, B. Arifaj, A. Lubishtani, F. Gashi, and E. Nishliu, "Aspect-based opinion mining of students' reviews on online courses," in *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*, 2020, pp. 510–514.

[90] R. Jha, A.-A. Jbara, V. Qazvinian, and D. R. Radev, "Nlp-driven citation analysis for scientometrics," *Natural Language Engineering*, vol. 23, no. 1, pp. 93–130, 2017.

[91] D. H. Widyantoro and I. Amin, "Citation sentence identification and classification for related work summarization," in *2014 International Conference on Advanced Computer Science and Information System.* IEEE, 2014, pp. 291–296.

[92] M. Bertin, I. Atanassova, C. R. Sugimoto, and V. Lariviere, "The linguistic patterns and rhetorical structure of citation context: an approach using n-grams," *Scientometrics*, vol. 109, no. 3, pp. 1417–1434, 2016.