Classifying Text in Citation Context as Relevant or Irrelevant to the Cited Paper

Afsheen Khalid[1*], Dilawar Khan[2], Shaukat Ali[3]
[1]Center for Excellence in IT, Institute of Management Sciences, Peshawar, Pakistan
[2]Computer Science &IT Department, University of Engineering and Technology, Peshawar, Pakistan, dilawarkhan@uetpeshawar.ed.pk
[3]Department of Computer Science, University of Peshawar, Peshawar, Pakistan, shoonikhan@uop.edu.pk
[*]**Correspondence:** Afsheen Khalid, afsheen.khalid@imsciences.edu.pk

Citation | Khalid. A, Khan. D, Ali. S, "Classifying Text in Citation Context as Relevant or Irrelevant to the Cited Paper", IJIST, Vol. 6 Issue. 3 pp 1088-1098, August 2024
DOI | https://doi.org/10.33411/ijist/20246310881098
Received | July 10, 2024 **Revised |** July 29, 2024 **Accepted |** July 31, 2024 **Published |** August 01, 2024.

Citation contexts, whether in the form of full citing sentences or text within a fixed window around the citation, have been widely used in various citation analysis applications. However, the absence of precise techniques to identify the exact span of text describing citations forces these applications to rely on extended texts as citation contexts. In this paper, we introduced new features combined with baseline features to accurately identify text that characterizes citations. Specifically, we utilized a Conditional Random Field (CRF) sequence classifier to categorize the surrounding text of citations as relevant or irrelevant. The integration of these features enhances the precision, recall, and F-measure scores for the Relevant (R) class. Although the average values of all measures are similar to those obtained with baseline features alone. Our approach significantly improves the extraction of relevant text.

Keywords: Citation Context; Conditional Random Field; Fixed Window; Citation Analysis; Relevant Text.

Introduction:

In scientific writing, citations summarize the key findings of referenced papers and serve as vital sources for knowledge accumulation. Additionally, the context in which a citation appears in a scientific article, known as the Citation Context (CC), which has gained significant importance over the last two decades. Citation contexts have been employed in various applications (e.g., [1][2][3][4][5][6][7][8]) either as the citing sentence with an explicit citation mark or as surrounding text within a fixed window. For instance, in sentiment analysis tasks, two sentences before and after the citing sentence are used as the CC [4][5]. These applications typically require only fragments of text or specific words or phrases directly related to a target citation, not the entire lengthy citation sentence(s). However, identifying only the relevant words in long sentences related to a target citation is challenging, compelling applications to utilize the entire text in the CC. To the best of our knowledge, there is no existing technique or algorithm that can extract the relevant text of a target citation from a CC effectively. Consequently, the presence of unrelated text within a CC can hinder the efficiency of tasks dependent on citation contexts. Ideally, only text fragments that characterize the target citation should be included in a CC.

In this paper, we aim to identify the words in lengthy CCs that directly describe the target citation. Below are examples of CCs, where the text in bold denotes the required relevant text related to the target citation, which is marked as key Citation. For example, combining our system with a fast sentence alignment program such as that of (key Citation), which performs alignment at a rate of 1,000 sentences, would make it possible to rapidly and accurately create a bilingual aligned corpus from raw parallel texts. Probabilities are based on relative frequencies, which are derived from the measures which are defined in (key Citation) as baseline.

These examples illustrate that only a few words in a lengthy CC are relevant, yet applications using CCs include all words, whether relevant or not. Our objective in this paper is to address the challenges of identifying relevant words in a CC. Using only the relevant text from CCs significantly affects applications in the citation analysis field. For instance, identifying relevant papers can be enhanced by improving the link information connecting two articles in a citation network. Similarly, in creating summaries based on CCs and improving information retrieval results, text that solely describes the citation is more useful. In the realm of generating summaries of citing papers [3][9][10], researchers have traditionally focused on selecting one or more complete sentences for a target citation. Unlike other studies, [10] presented approaches for extracting relevant spans of the cited paper to create accurate summaries. However, their dataset had only CCs with multiple references, where citation marks act as delimiters, making it less challenging to define reference text compared to CCs with a single citation.

The concept of using CCs in information retrieval was introduced by [11], who demonstrated that index terms from CCs significantly impact search results. They compared various window sizes to determine optimal reference terms, with a window size of fifty yielding the best results. Despite the adoption of this window size by many researchers, context window size detection remains an unresolved problem, as noted by [12]. The current paper introduced new features to be utilized along with the CRF machine learning model for identification of relevant text within citation contexts. We incorporated similar features as described in [10] along with new ones derived from our previous work [13]. We classified each word in the CC as Relevant (R) or Irrelevant (I). Additionally, we compared our results with those of [10], demonstrating that our approach yields higher accuracy with datasets containing single or multiple citations in the CCs.

Objectives:

- Identifying the words which directly describe a citation in sentences with single citation or in sentences with multiple citations**.**

- Designing general features for sequence classifiers to classify the text in CCs as relevant or non-relevant
- Checking improvement in results by using DP tree features.
- Checking which features of DP tree perform better.

The rest of the paper is organized as follows: Section 2 describes related work. The methodology of this study is outlined in Section 3. Classification results, along with discussion, are presented in Section 4. Finally, Section 5 provides the conclusion and future directions of our research.

Related Work:

There has been extensive research on identifying relevant and non-relevant text for citations in CCs. This task is commonly addressed in the fields of scientific article summarization and information retrieval. We have highlighted some recent work related to this problem. The concept of using reference terms in information retrieval dates back to [11], where the authors attempted to find ideal reference terms from CCs but lacked the techniques to automatically extract these terms. Liu et al. [14] took a different approach by retrieving CCs based on query terms for literature retrieval. These query terms, referred to as reference terms, appeared in all the retrieved CCs and were used as citation topic information. However, they did not verify if these terms were used to describe a citation or if they were noisy terms. Another study in information retrieval [1] introduced exploratory search using CCs, employing 100 words on either side of the citation mention for automatic query refinement. Similarly, [2] used 400 characters around the citation point to rank documents based on CCs, without distinguishing whether the words characterized the target citation.

Jha et al. [10] addressed the issue of citing sentences containing multiple references, which altered reference scopes for different citations when summarizing scientific papers. They used various machine learning algorithms to identify the reference scope of citations and found that the CRF model outperformed others. Cohan et al. [3] tackled the problem of inaccurate citation texts by using the cited paper to find relevant text. They performed contextualization as an information retrieval task, extracting textual spans from the cited paper and indexing them with an IR model. The citation context from the citing paper was used as a query to obtain more relevant context from the indexed spans.

In sentiment analysis, [5] used complex linguistic patterns to classify citation text fragments into positive, negative, or neutral classes. They utilized four sentences: one with explicit citation marks, two preceding, and one following it. Another author [4] also experimented with detecting the sentiment of citing authors towards cited articles, showing that deep learning performed better for larger samples, while support vector machines were more effective for smaller samples. They demonstrated that context-based samples were more beneficial than context-less ones for citation sentiment analysis. Ghosh et al. [15] classified ACL Anthology papers into three sentiment classes, considering only the citation sentence. They emphasized the importance of sentences before and after the citation sentence. However, due to the lack of techniques to properly recognize the text span relating to the citation, researchers often use only the sentences containing the citation. For sentiment analysis tasks involving longer CCs, more precise results can be achieved by first extracting the relevant words of the target citation.

Recent research [16] argued against using the symmetric window method for citation recommendation tasks. Instead, they showed that sentence-based approaches were superior, using manual annotation of the ACL Anthology Network (AAN) dataset to demonstrate that a five-sentence context outperformed symmetric window methods. Kang et al. [17] manually analyzed numerous citing sentences to determine the characteristics of citing sentences and citation scope. Their main observation was that only 5% of citing sentences were multi-sentence

CCs. Due to the lower ratio of multi-sentence CCs, most research in this area focuses on single citation sentences. Classification of citation functions is crucial in citation analysis, with studies [6][7][8] focusing on the text near citations. As a result, single citing sentences are often used in citation function taxonomies. Citation is important for classification, another key task, was explored by [18], who aimed to identify features with high predictive value. This type of study requires the full text of the publication.

An earlier study [19] developed a classifier for the classification of citations into two classes: important and non-important. Important citations are considered as those which either use the cited work or extend it in any way. Non-important citations are taken as those citations which appear in the related work section of a scientific document or those which are used for the comparison of some earlier work. This type of work is the citation function classification task, which is quite different from our work. The set of features used in this work is also non-identical to the features used in our work. However, in this study, noun phrases before the citation and a verb and noun after it were used in building one of the features, but authors acknowledged that some noun phrases did not describe the citations and emphasized the need for robust heuristics to accurately capture relevant nouns and verbs. Moreover, we cannot compare the results of this work with the problem of our paper due to dis-similarity of nature of work.

Material and Methods:

To address the problem of identifying relevant reference text in CCs, we employed a supervised machine learning approach using CRF for classification. CRF is a probabilistic framework ideal for structured prediction tasks. Our choice of CRF is inspired by Amjad et al. [10], who demonstrated its superior performance in similar tasks. Unlike Amjad et al., who focused on sentences citing in multiple papers, we considered all types of CCs, whether the author of [10] cited single or multiple references. This broad approach accommodates the varied linguistic structures and lengths of single citation sentences and the often entity-focused multiple citation sentences.

The workflow of our methodology started with preparing the dataset of 113 CCs articles obtained from ANN dataset, preprocessing it and then annotating it for pointing out the text which describes a citation. The annotated dataset was used to extract various features which are utilized to train the CRF classifier. Furthermore, the annotated dataset was divided into training and test datasets. The classifier was trained on the training dataset with the extracted features. At the end, the trained classifier was tested on the test dataset and finally the results were compared with the baseline work. Figure 1 depicts all these steps in pictorial form.

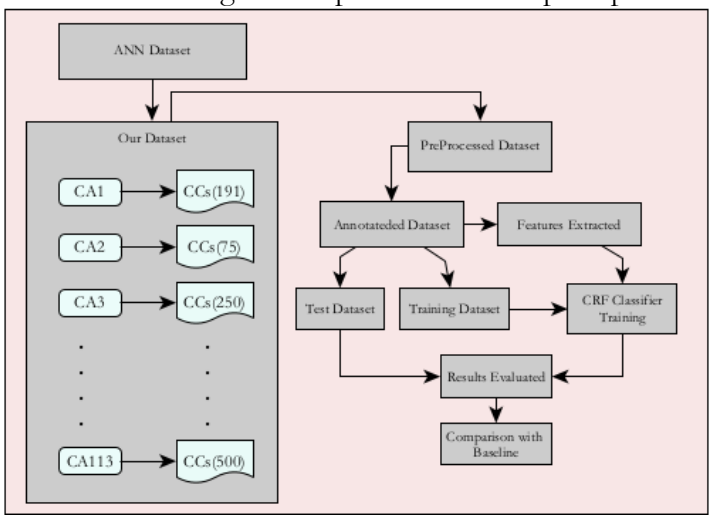

Figure 1. Flow diagram of methodology. CA represents a Cited Article and CCs means Citation Contexts.

Dataset:

Due to the absence of standard CC datasets, we compiled our dataset from the AAN interface, focusing on computational linguistics and NLP. We excluded other fields to avoid unfamiliar terminology. Our dataset consists of 113 articles from 1980 to 2020 with varying numbers of CCs. The maximum number of CCs of an article is 604 and minimum number is 30 CCs in our dataset. Other than these, many articles have total CCs between 100 and 200 range, and only a few articles have CCs between 300 and 500 range. After preprocessing, our dataset contains a total of 9970 citation sentences. These sentences were annotated by two human experts, achieving an inter-annotator agreement of 0.60 which indicates a substantial agreement.

Preprocessing:

The text downloaded from ANN interface was not in the condition to be used directly by machine learning model to get reliable results. For cleaning the Text, we manually searched for and removed the unnecessary text, if there were no specific patterns that could be addressed by a script. As scientific documents focus on different topics, in some instances, the correction of terminology and removal of mathematics and unusual symbol sequences from these scientific documents was a time consuming and hectic task.

To prepare the text for the CRF model, we utilized several Python scripts:

- **Cleaning**: Removed mathematical expressions and non-English text.
- **Citation Filtering**: Removed non-focus paper citations and replaced key paper citations with "key Citation."
- **Punctuation and Case Normalization**: Removed punctuation and converted text to lowercase.
- **POS Tagging and Dependency Parsing**: Generated Part-of-Speech (POS) tags and dependency parse trees using the nltk and spaCy libraries.

Classes:

For text classification within CCs, we employed two classes: R & I. The class "R" was assigned to words pertinent to the citation, while "I" was used for words outside the reference scope. We developed a Python script to label words based on reference words provided in Excel files. Words matching the reference were labeled as "R", and those not matching were labeled as "I". This process prepared our dataset for the classification algorithm. We utilized the CRF suite library of Python [20] for CRF classification. Our dataset comprised 7,970 CCs for training and 2,000 CCs for testing. During training, we employed 10-fold cross-validation to enhance accuracy.

It is important to note that our dataset does not suffer from class imbalance, which is a common issue in classification tasks. This balance is due to the nature of citation sentences: when the citation marker appears at the end of a sentence, most words are labeled "R", and few are "I". Conversely, when the citation is at the beginning or middle, few words are "R" and many are "I". This variation ensures that the overall distribution of classes remains balanced across the dataset, mitigating imbalance issues and preserving the performance of the CRF.

Sequence Classification:

Our task requires considering the impact of preceding and subsequent words on the classification of a word. Words often inherit the label of their surrounding words if they are part of the same named entity or clause. Therefore, rather than classifying words in isolation, we use a sequence classifier. This method incorporates features from both the preceding and following words, in addition to the word being classified. Unlike previous studies, which considered only the features of the target word, our approach also includes features from neighboring words. We leveraged dependency relations from a dependency parse tree, inspired by previous work [13]. While that work used an algorithmic approach to extract relevant citation text, it provided valuable insights and features for machine learning models. We adapted these insights to

enhance our feature set for sequence classification. We experimented with a variety of features, including those from previous studies and new ones, detailed as described in Table 1.

Table 1. Features used in CRF classifier and their description.

Number	Feature	Description
1	Is-stop-word	Is the word in the stop word list of the English language. Stop words are words like is, of, to, etc.
2	Part-of-speech (POS)	Coarse-grained or simple part-of-speech like Noun, Verb, etc.
3	POS-tag	Fine-grained or detailed part-of-speech like VBZ, VBG, or NNP, etc.
4	Dependency-relation	Syntactic dependency relation of the word in DPT.
5	Left-token-POS	The POS of the leftmost word of the syntactic descendants.
6	Left-token-POS tag	The POS tag of the leftmost word of the syntactic descendants.
7	Right-token-POS	The POS of the rightmost word of this word's syntactic descendants.
8	Right-token-POS tag	The POS tag of the rightmost word of this token's syntactic descendants.
9	Head-POS	POS of the syntactic parent of the word.
10	Head-POS-tag	POS tag of the syntactic parent of the word.
11	Head-dependency-link	Syntactic dependency relation of the parent of the word in DPT.
12	Distance	Distance between the key citation and the word.
13	Place	The word is before or after the key citation.
14	Is Noun	The POS of the word is NOUN.
15	Shortest-path	The length of the shortest dependency path between the word and key citation in DPT.
16	Is-in-path	Is the word in a path from the root to key citation in DPT and satisfies the Is-correct-obj feature.
17	Is-ancestor	Is there a common node between the ancestors of the key citation and the parent of the word.
18	Is-correct-dobj	The POS tag of the word is dobj and its parent's POS is VERB.
19	Is-it-child-dobj	If any of the ancestor of the word is correct dobj according to the Is-correct-dobj feature.
20	Is-correct-nsubj	The POS tag of the word is nsubj and its POS is not PRON.
21	Is-it-child-nsubj	If any of the ancestors of the word is correct nsubj according to the Is-correct-nsubj feature.
22	Is-comp	Is the dependency relation of the word ccomp, xcomp, or compound.
23	Is-in-other-paths	Does the word belong to the paths acl-agent-pobj, acl-prep-pobj or pobj-prep-pobj of DPT.

Features and Implementation:

Features 3, 4, 6, 8, 12, 13, and 15, which are described in Table 1, are also previously employed by [10]. For the preceding and subsequent words, we utilized all 23 features mentioned in the table, modifying the feature names to distinguish between them. We prepended "-1:" for the previous word's features and "+1:" for the next word's features, resulting in a total of 69 features associated with each word. The impact of considering only the 23 features of a single word versus including features of the preceding and subsequent words is discussed in the next chapter.

Evaluation Measures:

To evaluate our results, we used standard measures of precision, recall, and F-measure. We calculated micro precision, micro recall, and micro-F-measure for each of the two classes and then averaged them to obtain macro values. These measures are detailed below:

Micro-Precision:

This feature measures the proportion of correct predictions of the relevant class R for a single cited paper. It calculates the proportion of CCs that belong to the positive class relevant R out of all predictions for class R, whether positive or negative.

$$\text{precision}_{\text{micro}} = \frac{TP}{TP+FP} \quad (1)$$

Micro-Recall:

It tells the ratio of CCs providing the correct positive class relevant R in total CCs of a single cited reference.

$$\text{recall}_{\text{micro}} = \frac{TP}{TP+FN} \quad (2)$$

Micro-F:

This metric measures the accuracy of the system using the harmonic mean of micro-precision and micro-recall. It checks how similar the precision and recall values are for a particular cited reference and hence finds how well the system performs. It helps in minimizing the influence of large precision and recall values.

$$F_{\text{macro}} = \frac{2*\text{precision}-\text{macro}*\text{ recall macro}}{\text{precisionmacro}+\text{ recall macro}} \quad (3)$$

Macro-Precision:

For all the target papers, this measure quantifies the average of the micro-precision values.

$$\text{precision}_{\text{ macro}} = \frac{1}{n}\sum_{i=1}^{n} \text{precision}(pi) \quad (4)$$

Macro-Recall:

It provides the average of micro-recall values for all the target papers in our dataset

$$\text{recall}_{\text{ macro}} = \frac{1}{n}\sum_{i=1}^{n} \text{recall}(pi) \quad (5)$$

Macro-F:

It measures the average of micro-F values for all the target references in our dataset

$$F_{\text{macro}} = \frac{1}{n}\sum_{i=1}^{n} F(pi) \quad (6)$$

Discussion and Comparison of Results:

We initiated our experiments by incorporating the features from [10] in our dataset. These features provided the highest performance with the CRF classifier, outperforming SVM and logistic regression classifiers. Based on these results, we selected a CRF classifier for our problem and used it for comparison. We reiterated the significance of Amjad et al.'s research in the next paragraph to establish it as a baseline for comparison. Notably, this research is the most relevant to our task, directly aligning with our objectives. Therefore, we used their results as a benchmark for comparison.

The dataset used by [10] comprises sentences that cite multiple references. Each instance of a CC has at least two separate citation mentions. In this scenario, the reference text of the key citation usually consists of one grammatical fragment, delimited between the two references or positioned at one end of the sentence. The presence of other references provides a useful marker to separate the relevant and irrelevant text of the key citation. This behavior is reflected in the features used in their work. The training/testing set in this research comprises 3300 citing sentences. The Avg-DS2 row of Table 2 shows the average values of precision, recall, and F-measure of the CRF classifier for the features used in the baseline. We conducted a series of experiments varying the number of features used in our classification task. For all experiments, we utilized the training and test datasets comprising 9970 CCs as described in methodology section.

Table 2. Results of CRF Classifier for Baseline and Added Features. Avg-DS1: Average scores of our dataset with baseline features. Avg-DS2: Average scores of the baseline dataset.

			Precision	Recall	F-measure
Baseline features	**Avg-DS1**	R class	**53.4** %	**50.7** %	**52.0** %
		I class	84.9 %	85.4 %	85.1 %
			79.9 %	**80.0 %**	**79.9 %**
	Avg-DS2		80.1 %	94.2 %	86.6 %
Added features	**Average**	R class	74.1 %	78.1 %	76.1 %
		I class	89.1 %	86.7 %	87.9 %
			84.2 %	**83.9 %**	**84.9 %**

It is important to note that our dataset is three times larger than the baseline dataset. Initially, we trained the classifier using the features from the baseline study and evaluated it on our test set. Comparison of the results for baseline and our added features is tabulated in Table 2. A comparison of our added features with the baseline is shown in Figure. 1.

The first three rows of Table 2 present the results for the R class, I class, and their average scores on our dataset which is referred to as Avg-DS1. The average classification scores on our dataset, which includes a diverse range of language patterns, are comparable to the baseline scores.

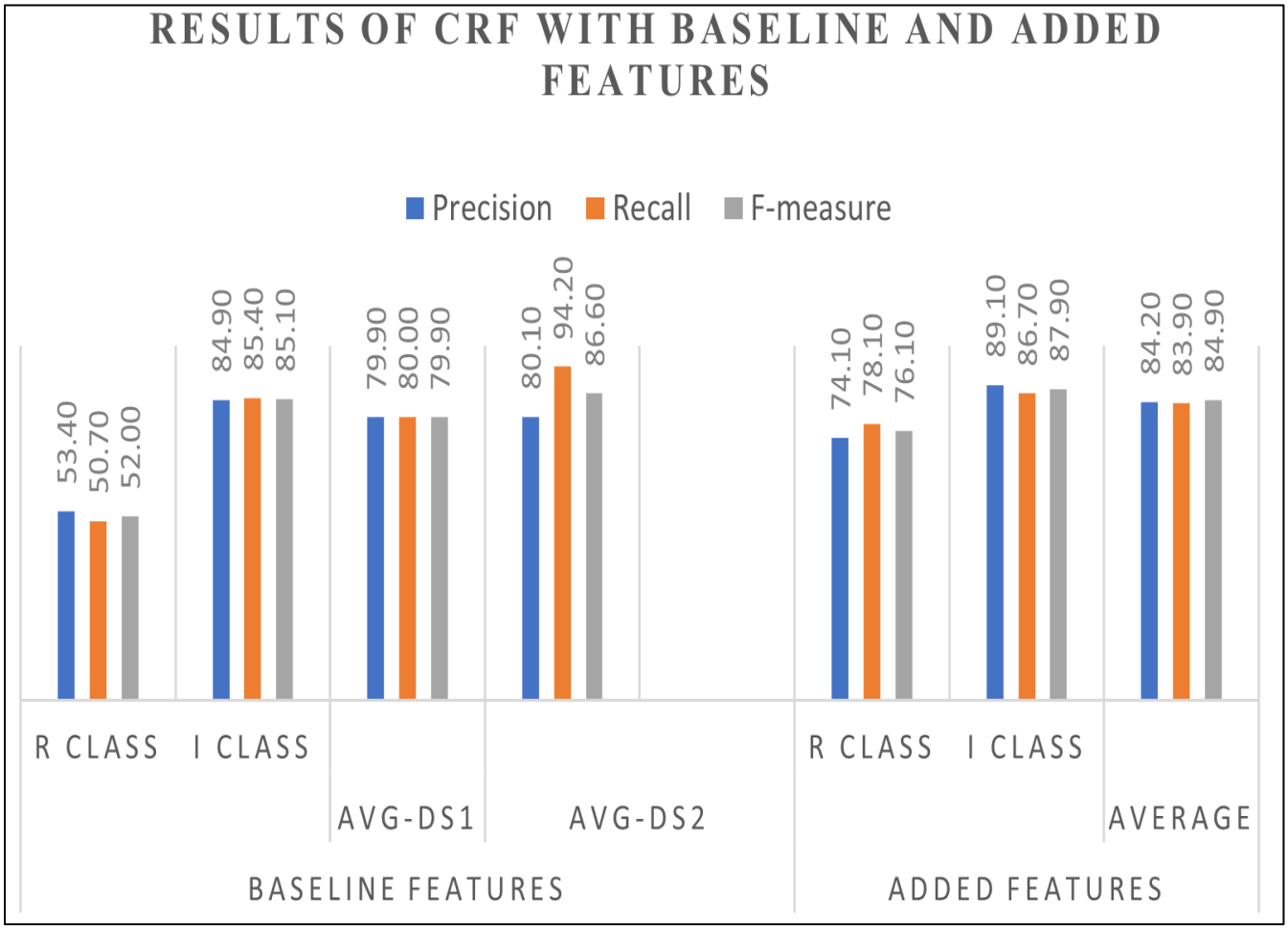

Figure 2. Results of CRF classifier. Avg-DS1 are the average scores of our dataset for baseline features. Avg-DS2 represents the average scores of the baseline dataset. The right-hand side

average represents the results of classification with added features.

However, the results in the first row indicate that all evaluation metrics for the class R is lower as compared to the class I. Since our primary focus is on improving the performance for class R, these scores are less satisfactory and require enhancement. To address this, we incorporated additional features. Specifically, we included features from one word before and one word after the target word, in addition to the features of the word being classified. This addition led to improvements in precision, recall, and F-measure for the R class, as illustrated in Figure. 3. Although the average values for all metrics remained similar to those achieved with only the baseline features, the inclusion of these additional features has enhanced the extraction of relevant text to a significant extent.

Furthermore, our experiments consistently showed higher scores for theI class compared to the class R. This disparity arises because the irrelevant class typically encompasses a larger portion of the text. We observed that by employing the dependency parsing algorithm [13], we gained better control over identifying relevant text. This approach ensures that even if some relevant words are missed or not removed, the output remains comprehensive, thereby improving the true positive counts. Moreover, we found that converting complex heuristics into features for machine learning is not always feasible.

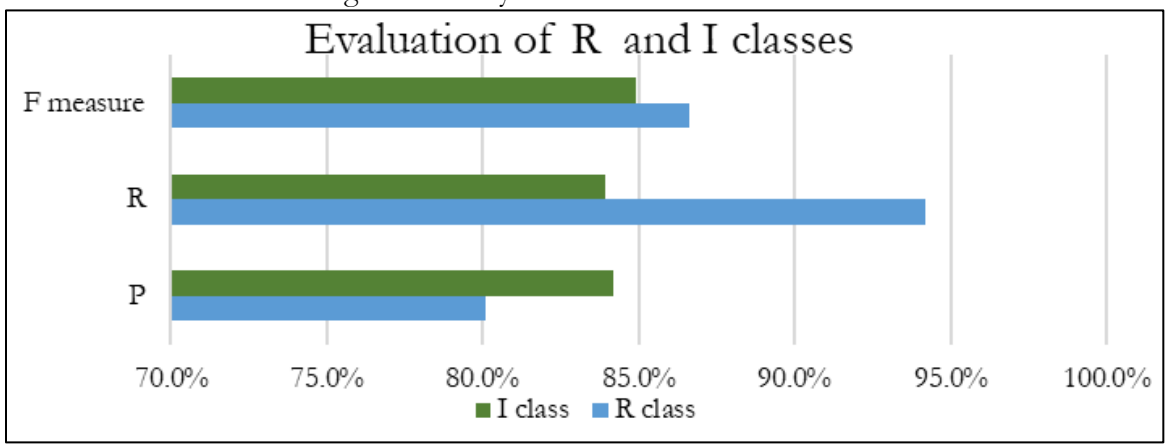

Figure 3. Evaluation measures P, R and F measure for R and I classes.

It is to be noted that the features of a word and the features of one word before and after it have a significant effect on the performance of the classifier. Along with this, the simple features of DPT have also been more effective than the features which represent lengthy paths in the DPT. Although the addition of long paths DPT features solve the problem of long-distance dependencies of a focus word on distant words in the citation sentence, such scenarios are not very frequent and common in citation sentences. We noticed that the long path DPT features added computational complexity without significantly impacting the performance of the classifier.

Conclusion:

We aimed to design features that align with the heuristics derived from the dependency parse tree. By utilizing a CRF sequence classifier and distinguishing between relevant and irrelevant classes, we achieved overall accuracy comparable to the baseline method. However, the baseline features yielded lower accuracy for specific evaluation measures. Our newly added features enhanced the classifier's accuracy for the relevant class by up to 30%. Future work will focus on leveraging deep learning models to further refine the identification of relevant citation text.

Author's Contribution: The first two authors contributed equally at each step of this paper till revisions. The third author helped in reviewing and incorporating the suggestions of the reviewers.

Conflict of Interest: There exists no conflict of interest for publishing this manuscript in IJIST.

Reference:

[1] R. Moro, M. Vangel, and M. Bielikova, "Identification of Navigation Lead Candidates Using Citation and Co-Citation Analysis," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9587, pp. 556–568, 2016, doi: 10.1007/978-3-662-49192-8_45.

[2] "Content sensitive document ranking method by analyzing the citation contexts." Accessed: Aug. 03, 2024. [Online]. Available: https://patentimages.storage.googleapis.com/ad/09/9c/3adaea688a2e2d/WO2016099422A2.pdf

[3] A. Cohan and N. Goharian, "Scientific document summarization via citation contextualization and scientific discourse," Int. J. Digit. Libr., vol. 19, no. 2–3, pp. 287–303, Sep. 2018, doi: 10.1007/S00799-017-0216-8/METRICS.

[4] K. Ravi, S. Setlur, V. Ravi, and V. Govindaraju, "Article Citation Sentiment Analysis Using Deep Learning," Proc. 2018 IEEE 17th Int. Conf. Cogn. Informatics Cogn. Comput. ICCI*CC 2018, pp. 78–85, Oct. 2018, doi: 10.1109/ICCI-CC.2018.8482054.

[5] "Sentiment classification based on linguistic patterns in citation context." Accessed: Aug. 03, 2024. [Online]. Available: https://www.currentscience.ac.in/Volumes/117/04/0606.pdf

[6] X. Su, A. Prasad, M. Y. Kan, and K. Sugiyama, "Neural multi-task learning for citation function and provenance," Proc. ACM/IEEE Jt. Conf. Digit. Libr., vol. 2019-June, pp. 394–395, Jun. 2019, doi: 10.1109/JCDL.2019.00122.

[7] M. Hernández-Alvarez, J. M. Gomez Soriano, and P. Martínez-Barco, "Citation function, polarity and influence classification," Nat. Lang. Eng., vol. 23, no. 4, pp. 561–588, Jul. 2017, doi: 10.1017/S1351324916000346.

[8] C. Martinez-Perez, C. Alvarez-Peregrina, C. Villa-Collar, and M. Á. Sánchez-Tena, "Current State and Future Trends: A Citation Network Analysis of the Academic Performance Field," Int. J. Environ. Res. Public Heal. 2020, Vol. 17, Page 5352, vol. 17, no. 15, p. 5352, Jul. 2020, doi: 10.3390/IJERPH17155352.

[9] N. Saini, S. Kumar, S. Saha, and P. Bhattacharyya, "Scientific document summarization using citation context and multi-objective optimization," Proc. - Int. Conf. Pattern Recognit., pp. 4290–4295, 2020, doi: 10.1109/ICPR48806.2021.9412201.

[10] R. Jha, A. A. Jbara, V. Qazvinian, and D. R. Radev, "NLP-driven citation analysis for scientometrics," Nat. Lang. Eng., vol. 23, no. 1, pp. 93–130, Jan. 2017, doi: 10.1017/S1351324915000443.

[11] A. Ritchie, S. Robertson, and S. Teufel, "Comparing citation contexts for information retrieval," Int. Conf. Inf. Knowl. Manag. Proc., pp. 213–222, 2008, doi: 10.1145/1458082.1458113.

[12] "Concit-corpus context citation analysis to learn function, polarity and influence", [Online]. Available: https://dialnet.unirioja.es/servlet/tesis?codigo=61958

[13] A. Khalid, F. Alam, and I. Ahmed, "Extracting reference text from citation contexts," Cluster Comput., vol. 21, no. 1, pp. 605–622, Mar. 2018, doi: 10.1007/S10586-017-0954-9/METRICS.

[14] 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, Oct. 2014, doi: 10.1007/S11192-014-1233-7/METRICS.

[15] S. Ghosh and C. Shah, "Identifying Citation Sentiment and its Influence while Indexing Scientific Papers," Proc. Annu. Hawaii Int. Conf. Syst. Sci., vol. 2020-January, pp. 2517–2526, Jan. 2020, doi: 10.24251/HICSS.2020.307.

[16] D. Duma, C. Sutton, and E. Klein, "Context matters: Towards extracting a citation's context using linguistic features," Proc. ACM/IEEE Jt. Conf. Digit. Libr., vol. 2016-

September, pp. 201–202, Sep. 2016, doi: 10.1145/2910896.2925431.
[17] I. S. Kang and B. K. Kim, "Characteristics of Citation Scopes: A Preliminary Study to Detect Citing Sentences," Commun. Comput. Inf. Sci., vol. 352 CCIS, pp. 80–85, 2012, doi: 10.1007/978-3-642-35603-2_11.
[18] D. Pride and P. Knoth, "Incidental or Influential? - Challenges in Automatically Detecting Citation Importance Using Publication Full Texts," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10450 LNCS, pp. 572–578, 2017, doi: 10.1007/978-3-319-67008-9_48.
[19] M. Valenzuela, V. Ha, and O. Etzioni, "Identifying Meaningful Citations," 2015, Accessed: Aug. 03, 2024. [Online]. Available: http://opennlp.apache.org
[20] M. Korobov, "Scikit-learn inspired api for crfsuite," 2016, [Online]. Available: https://github.com/ TeamHG-Memex/sklearn-crfsuite

Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.