

Article

Towards Potential Content-Based Features Evaluation to Tackle Meaningful Citations

Faiza Qayyum ¹, Harun Jamil ², Faisal Jamil ¹  and Do-Hyeun Kim ^{1,3,*}

¹ Department of Computer Engineering, Jeju National University, Jeju-si 63243, Korea; faizaqayyum@jejunu.ac.kr (F.Q.); faisal@jejunu.ac.kr (F.J.)

² Department of Electronic Engineering, Jeju National University, Jeju-si 63243, Korea; harunjamil@hotmail.com

³ Research Center of Advanced Technology, Jeju National University, Jeju-si 63243, Korea

* Correspondence: kimdh@jejunu.ac.kr

Abstract: The scientific community has presented various citation classification models to refute the concept of pure quantitative citation analysis systems wherein all citations are treated equally. However, a small number of benchmark datasets exist, which makes the asymmetric citation data-driven modeling quite complex. These models classify citations for varying reasons, mostly harnessing metadata and content-based features derived from research papers. Presently, researchers are more inclined toward binary citation classification with the belief that exploiting the datasets of incomplete nature in the best possible way is adequate to address the issue. We argue that contemporary ML citation classification models overlook essential aspects while selecting the appropriate features that hinder elutriating the asymmetric citation data. This study presents a novel binary citation classification model exploiting a list of potential natural language processing (NLP) based features. Machine learning classifiers, including SVM, KLR, and RF, are harnessed to classify citations into important and non-important classes. The evaluation is performed using two benchmark data sets containing a corpus of around 953 paper-citation pairs annotated by the citing authors and domain experts. The study outcomes exhibit that the proposed model outperformed the contemporary approaches by attaining a precision of 0.88.

Keywords: citation analysis; citation classification; binary classification; random forest; kernel logistic regression



Citation: Qayyum, F.; Jamil, H.; Jamil, F.; Kim, D.-H. Towards Potential Content-Based Features Evaluation to Tackle Meaningful Citations. *Symmetry* **2021**, *13*, 1973. <https://doi.org/10.3390/sym13101973>

Academic Editor: Sergei D. Odintsov

Received: 9 September 2021

Accepted: 15 October 2021

Published: 19 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A scientific study is usually conducted by capitalizing on the earlier researches of peers in a domain. It establishes a connection with precursory studies via “citation”. A citation serves as an acknowledgment that a document receives from another paper in reciprocation of referring the study [1,2]. Besides this, the citation has a crucial role in forming decisions for multifarious academic policies, such as research grant allocations [3], clustering of publications, peer judgment [4], authors ranking [5–8], assessing the academic influence of country [9], in diversified disciplines ranging from machine learning [10], Internet of things (IoT), networking, etc. [11–13].

These policies primarily utilize quantitative citation analysis-based measures wherein a mere count of a citation is considered. A high count of a citation is deemed as an indicator to correlate with the prestige of a publication, author, institute, etc. [8].

Each citation reason serves a different purpose, thus varying significance, which discourages treating all citations equally. The quantitative citation analysis-based approaches assign equal weight to all the citations irrespective of the reason a particular citation has been made [14–17]. The scientific community argues against harnessing pure quantitative citation analysis-based measures and argues that the reason for a citation must be contemplated [16–18]. The majority of the researchers sift through misleading citations prior to employing them to measure the policies mentioned above [17,18]. Back in the late 90s, citing

authors were interviewed to provide the reason at the time of publication [19,20]. However, the method did not garner approval as it involves a complex manual process. After that, researchers have proposed different citation classification methodologies that manually scrutinize research papers' content to determine their classes [21,22]. Finney floated the idea that the process may be automated by capturing clues from research papers [23]. Her idea was substantiated by [24] in the form of a first fully automated citation classification technique that considers cue-terms and linguistic features for the classification. However, the study has been criticized due to overlapping and large categories (i.e., 35). Afterward, various other approaches have also been presented to classify citations into a varying number of reasons. There has been a continuous dispute regarding the sufficient number of citation classes to serve the objective of refining citation count-based measures [19,24–26].

Moreover, one of the critical issues faced by the citation classification community is incomplete data. Typically, data corpus of such a nature involves symmetric and asymmetric data. In the context of the scenario being considered in this study, there are few benchmark datasets, making the asymmetric citation data-driven modeling quite complex. The missing parts of data sets pose a challenge in exploiting the available incomplete data in the best possible way so that the accurate information may be ascertained with significant accuracy. Presently, the scientific community is more inclined towards reducing the number of categories into two (i.e., important and incidental) to tackle only meaningful reasons by exploiting the contemporary incomplete data sets of unstructured or semi-structured nature appropriately to discover the hidden knowledge pertaining to the accurate class of a citation. This idea has been implemented in binary citation classification wherein citations are classified into (1) important and (2) incidental categories [2,10,14,16,17]. Similar to the aforementioned studies, we consider that classification of citation into the said two categories would play an immense role in finding meaningful citations. Valenzuela et al. are at the vanguard of classifying citations into important and incidental categories by using metadata and content-based features with SVM and Random Forest classifiers [16].

Now a question crops up that which citation reason should be considered important or incidental? The existing classic binary citation classification considers those citations as important which inspire the citing study for using or extending it, whereas, incidental citations contribute to the citing work in terms of explaining the background theme of a study [2,14,16,17,27].

Besides refining the quantitative citation analysis-based approaches, the binary citation classification can also help to find highly relevant research material for researchers. For instance, consider the following scenario: researchers pursuing research degrees in different disciplines, pose a query on the web to find closely relevant research documents against the research topic being focused on. Web sources return millions of records exhibiting them as relevant papers wherein only a few are actually relevant. On the other hand, if citations of the papers related to the focused topic are considered and further classified into important and incidental, then there is a high probability that the user will have a maximum number of relevant documents, unlike the existing web sources.

The contemporary classification studies exploit different features relating to metadata or the content of research papers [18,24–26]. The content-based features are dominant among others due to being rich in terms of meaningfulness. However, critical analysis of contemporary approaches depicts a need to incorporate some important aspects while selecting appropriate features. This study proposes a comprehensive methodology that exploits a list of novel content-based features to classify citations into important and incidental classes. The features include section-wise citation count, citation sentences, content similarity as a whole, and between Introduction, Methods, Results, and Discussion (IMRaD) sections of research papers. Another contribution of this study is to assess the potential of different parts-of-speech (PoS) terms appearing in citation sentences and also in IMRaD of research papers. Two benchmark datasets have been employed to evaluate the proposed study. The binary citation classification is performed using support vector machine (SVM), random forest (RF), and kernel logistic regression (KLR) classifiers. The

outcomes revealed that the proposed approach outperformed existing studies [2,16,17] by achieving precisions of 0.88 and 0.80 for Valenzuela's and Qayyum's data sets, respectively.

The rest of the paper is organized as follows: Section II presents related work; Section III deals with the proposed methodology. Finally, the study outcomes are presented in Section IV, and Section V concludes the paper.

2. Literature Review

The main idea to discover citation reasons was presented by [15]. The author identified fifteen reasons for citations. This study originated new dimensions of research towards finding other possible reasons. Subsequently, [21] presented thirteen other reasons for citations. These reasons were identified by analyzing 66 articles from multiple disciplines. The specified reasons pulled the scientific community towards the critical scrutiny of purely citation count-based approaches. Until now, various studies have contented regarding equal importance of citations. In 1975, Moravcsik and Murugesan presented the first manual technique by classifying the citations into four categories [28].

Nanba and Okumura [29] classified the citation into three types: (a) Type B: that states the relevance in terms of explaining methods and theories of other studies, (b) Type C: it states relevance in terms of comparing the related works or finding the existing issues and (c) Type O: these categories contain all those relations that do not fall in Type B and C. All of these schemes mentioned above have classified the citations by applying manual methodologies. The inclination towards automatic citation classification was increased after the idea presented by Finney [23]. Finney has automatically classified citations into five categories by employing cue phrases. Subsequently, Garzone and Mercer [23] proposed the first complete automatic citation classification scheme. Their system takes different articles as input and produces the corresponding citation function as an output. They have presented 35 classes for citations which were merged into ten categories. For classification, almost 200 linguistic rules were employed.

In 2003, Pham and Hoffman [24] harnessed cue phrases developed a rule-based knowledge system to classify citations into four categories. Teufel et al. [19] presented a citation classification model that segregates them into 12 classes, generalized into four types. This scheme has adopted rules from Spiegel's method [21]. In that study, the first time, the machine learning algorithm was utilized for citation classification. The scheme has attained an F-measure of 0.71 to conclude neutral category holds around 65% of the citations. Pride et al. [30] classified citations using features of [16] by changing the model's configuration settings. The study has been evaluated on the set containing 465 paper-citation pairs collected by [16]. This model has yielded a precision of 0.69. Another study proposed by Tandon et al. [31] harnessed citation context from research articles to produce its summary automatically. In this scheme, the citations are classified into five categories. A language model approach was employed for this classification in which the language models were developed for all five classes. The language models stipulating optimal probability to generate a certain citation context were deemed for the classification. The training set was formed using 500 citation contexts extracted from Microsoft Academy. The model resulted in 0.68 precision.

A binary citation classification model was recently presented by Valenzuela et al. [16] that classifies citations into two classes, i.e., Important and incidental. In this scheme, a dataset comprising of 465 (citing, cited) pairs were collected. Two domain experts performed the annotation of pairs. This was the first work in which citations were segregated into two classes. They proposed a novel machine learning model to classify the citations into binary categories using twelve features. This system has achieved the best result with the in-text citation count by obtaining an F-measure of 0.37.

Furthermore, while considering all the twelve features, the system has achieved 0.65 precision and 0.90 recall. In the same year, Zhu et al. [17] presented a binary (influential and non-influential) citation classification model. The authors have used five types of features and yielded that the "in-text citation count" feature outperformed other features.

Another study proposed by [27] performs binary citation classification by combining features of four state-of-the-art approaches [16,18]. The study reported 29 top-scored features with a precision of 0.89 for the data set containing 465 pairs collected by Valenzuela et al. In another study, authors [14] classified citations into important and incidental categories using metadata-based features. This study has been evaluated on the same two data sets which are employed in our proposed study. The study reported a precision of 0.72 attained using an RF classifier. Likewise, the study by [2] classified citations into important and incidental categories using features such as similarity score, IMRaD based, and overall citation count. Similar to our proposed technique, the study [14] uses two data sets: (1) Valenzuela et al.'s data set that comprises 465 pairs and (2) Qayyum et al.'s data set, which contains 465 pairs. The study [2] recently presented a binary citation classification model that uses KLR, SVM, and RF classifiers and formed features by computing sentiment analysis of in-text citations. The study has used the same benchmarks datasets as used by [14] and reported an F-measure of 0.83 and 0.67 for both datasets. Our proposed research presents a binary citation classification technique that primarily focuses on introducing a list of novel potential features that have not been given attention by the approaches stated above.

3. Methods

This section encompasses details about the systematic steps to classify citations into: (1) important and (2) incidental classes. A detailed architecture of the proposed system is shown in Figure 1. As explained earlier, this study primarily focuses on discovering the best features from the content of a research paper to maximize the accuracy of binary citation classification. We devise a comprehensive methodology that exploits a list of potential content-based features for binary classification. The employed features listed below are exploited in the best possible way as explained in the following sections:

- Potential of citation count in IMRaD
- Contribution of parts of speech (PoS) in citation context
- Contribution of bigram terms in citation context
- Presence of different parts of speech (PoS) in IMRaD
- The potential of overall and section-wise content similarity of pairs
- Combination of best-performing features

Two comprehensive data sets from [16] and [14,16] have been employed, and a list of potential features is extracted from them. Then, these features have been pre-processed to be prepared for the experimentation phase. After that, N-Gram, PoS Tagging, and semantic-based methods are applied to the features for their score calculation. A detailed explanation of all the applied methods is explained below.

3.1. Benchmark Dataset

Appropriate data plays a significant role in revealing various facts. Considering this aspect, we have employed two data sets that can help evaluate the proposed features in classifying citations into important and incidental categories.

3.1.1. Valenzuela's Dataset

The first data set has been collected by Valenzuela et al. [16]. The authors have acquired the annotations of citations as important and incidental for paper-citation pairs taken from a collection of 20,527 papers published in the domain of Information Sciences on ACL anthology. These papers contain around 106,509 citations. Valenzuela et al. [16] formed 465 paper-citation pairs and annotated them as important and incidental from two domain experts. This is the only data set of the required nature that is freely available online; we have chosen this to apply the proposed methodology. A total of 14.6% of pairs are important, and the remaining 85.4% of pairs are incidental among the pairs of this data set.

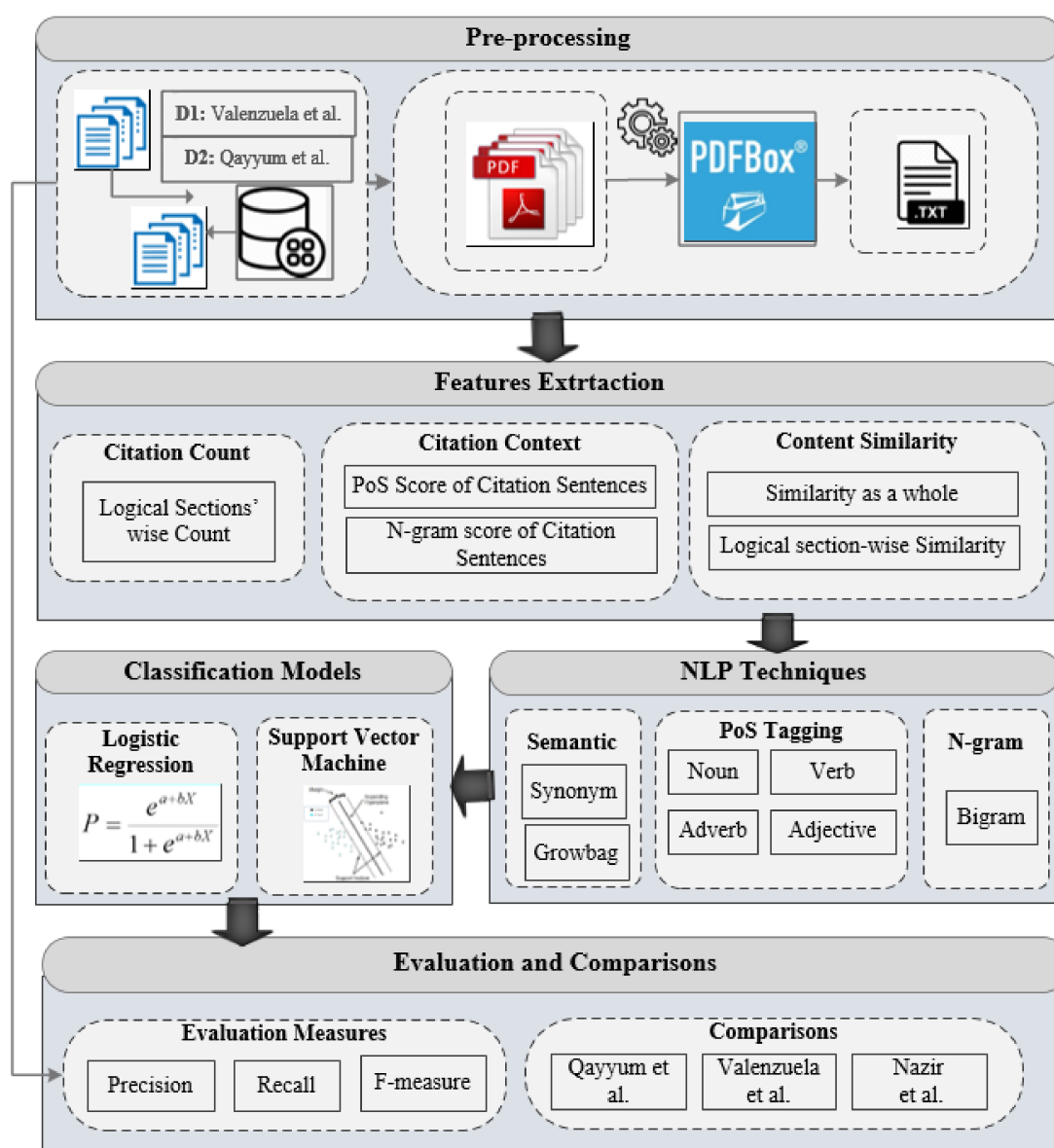


Figure 1. Systematic flow of proposed binary citation classification model.

3.1.2. Qayyum and Afzal's Dataset

Conclusions drawn from a single data set might not be adequate to assess overall results in the given scenario. For instance, 465 pairs are less in number, and there are only 14.6% important citations. Therefore, another data set was formed by considering the faculty members from Capital University of Science and Technology as citing authors and formed 488 paper-citation pairs. The data has been collected by one of our earlier studies [14]. These papers have been published in the domains of Databases, Information Science, Software testing, and networks. These pairs have been labeled as important and incidental by the citing authors themselves. The annotation formed 18.4% pairs from the important category.

3.1.3. PDF to Text Conversion

The authors [16] have only provided paper IDs of the annotated pairs published in the ACL anthology. We tracked those papers through their IDs and downloaded them. For the dataset by Qayyum et al. [14], we already have their portable document format (PDF) files as those were required to provide relevant materials to annotators to recall the citing reason. Since PDF files are hard to process and we require the text of the papers to apply

the proposed methodology, therefore, the PDF files were converted into (extensible markup language) XML using PDFX (portable document format exchange) tool. We extracted the required text using a script prepared for this purpose in python.

3.2. Features

As explained earlier, the proposed study primarily focuses on identifying potential features that have an essential role in discovering important citations. The features are extracted from the plain text of the pairs. The list of extracted features is shown in Table 1.

Table 1. Features description.

No.	Description
1	Section-wise citation count
2	Citation context: Bigram terms
3	Presence of noun in citation context
4	Presence of adjective in citation context
5	Presence of adverb in citation context
6	Presence of verb in citation context
7	Section-wise Similarity
8	Section-wise existence of noun
9	Section-wise existence of adjective
10	Section-wise appearance of verb
11	Section-wise appearance of adverb

3.2.1. Citation Count

A citation serves as a helpful measure in the decision-making of academic policies such as researchers' or institutions' ranking, funds allocation, finding cognoscenti in a domain, etc. For example, a research paper typically contains an Abstract, Introduction, Related Work, Methodology, Results, and Conclusion. In this study, we analyze the potential of citation count appearing in different logical sections. This has been carried out based on the following assumptions:

- Introduction and related work/literature review sections contain a comparatively higher number of citations [32]. We believe that these sections present a brief overview of the background knowledge of the topic or explanation of the terminologies in the domain. Hence, an author cites those studies that can connect with proposed research in terms of background knowledge (i.e., incidental citation).
- Methodology and results sections delineate information on the proposed methodology; therefore, it is highly likely to contain in-text citations of those papers that might have been extended or adopted by the proposed study.
- Based on the assumptions stated above, this study exploits in-text citations' existence in the IMRaD logical sections: Introduction, Methods, Results, and Discussion, using formulas 1, 2, 3, and 4, respectively. This count has been divided by the total count of in-text citations in the paper using Equation (1).

The following is the description of the formulas:

Let Sections = {I, M, R, D, F} where I represents "Introduction", M represents "Methodology", R represents "Results," D represents "Discussion," and F represents "Full-content".

- Consider the records shown in Figure 2 from D1. Each row in the above records represents a pair; thus, as per the Figure, there are 12 pairs (which are 465 in actuality for D1). Let i be the citing paper (shown in column A) and $i = \{1, 2, 3, \dots, n\}$ where n represents total number of citing papers. Let j be the cited papers (shown in column B), and $j = \{1, 2, 3, \dots, m_i\}$ be the count of total m number of cited papers. Since the number of cited papers for each citing paper would be different, therefore, we consider m_i to be the total count of citing paper as per i cited paper. In the context of the following Equations (1)–(4), let S be the citing paper and C be the cited paper. So, $S_i C_j$ would represent the i th citing paper and its corresponding j th cited paper.

	A	B
1	Citing Paper	Cited Paper
2	A00-1043	C00-2140
3	A00-1043	P02-1057
4	A97-1011	W09-1118
5	A97-1011	A00-2017
6	A97-1011	C00-2099
7	A97-1011	W04-1505
8	A97-1011	P99-1033
9	A97-1011	W06-0202
10	A97-1011	P01-1006
11	A97-1011	E12-1072
12	C00-1072	C02-1130
13	C00-1072	C04-1077

Figure 2. Overview of pairs from dataset.

Now, the following formula computes the number of times cited paper appears in the “Introduction” section of the citing paper. For instance, the numerator in Equation (1) is the total count of in-text citations in the introduction section of the j th cited paper in i th citing paper. The denominator represents the total number of in-text citations of j th cited paper in i th citing paper.

Let us assume a pair as (A, B), where A represents citing paper and B represents cited paper, which means “A” has cited “B” in its references, so obviously “B” must have been cited within the body of paper “A” which are termed as “in-text citations”. Let us suppose that “B” appears 8 times in the overall body of the paper and 4 times in the “Introduction” section of the citing paper, so the formula shown in Equation (1) will be computed as the ratio of 4/8. The same procedure is followed for all the remaining sections (i.e., Methods, Results, and Discussion). Similarly, Equations (2)–(4) computes the citation count of j th cited paper in i th citing paper in “Methods”, “Results” and “Discussion” sections. Equations (2)–(4), computes the citation count in the “Methods” section, respectively.

$$CC_{Introduction}(S_{ij}) = \frac{CC_I(S_i C_j)}{CC_F(S_i C_j)} \quad (1)$$

$$CC_{Methods}(S_{ij}) = \frac{CC_I(S_i C_j)}{CC_F(S_i C_j)} \quad (2)$$

$$CC_{Results}(S_{ij}) = \frac{CC_I(S_i C_j)}{CC_F(S_i C_j)} \quad (3)$$

$$CC_{Discussion}(S_{ij}) = \frac{CC_I(S_i C_j)}{CC_F(S_i C_j)} \quad (4)$$

3.2.2. Citation Context

A sentence containing an in-text citation is known as a citation sentence [16]. While citing a study in text, authors mention the description that can provide a clue regarding the purpose of a citation. The description comprises such words that can provide a vital indication regarding the reason for a citation. Consider the following two sentences as an example:

Sentence 1: “this study further investigates the problem addressed by [5]”

Sentence 2: “the study [6] also explains this theory”

The terms used in the first sentence such as “further”, “investigates”, “problem” hint that this citation belongs to the “important” category. On the contrary, terms appear-

ing in sentence 2, “explains”, “definition,” provide a clue that this citation is from the “incidental” category.

In this study, we have extracted such terms from citation sentences in two dimensions: (1) Unigram and Bigram terms (2) PoS including a verb, adverb, adjective, and adverb. Such terms have been extracted from 70% of the pairs used for training. The terms are maintained in the lexicon, verified by a domain expert from the domain Computer Science who has a strong command over the English language and can differentiate terms from important or incidental categories. The following are the steps performed to extract the terms.

A. Pre-processing:

This step is mandatorily performed in the scenarios wherein text is required to be processed. The pre-processing phase removes all the noise or redundant information from the data. In this study, the stop words were removed, and the terms were converted into root terms via stemming.

The detail is given below.

- **Stop Words Removal:** Different English words fail to provide any clue regarding relevance to the particular class. These words include “is”, “are”, “am”, “the” etc., and are known as stop words. We have removed the stop words from extracted citation sentences using Onix Text Retrieval Toolkit.
- **Stemming:** Stemming is performed to convert terms into their base terms so that there is no need to keep a separate record for semantically similar terms. We have used the porter stemmer algorithm [33] to stem the terms of citation sentences. For example, stemming converts terms such as “computing, computer and computes into comput”, etc.

B. Bigram Score:

Analyzing a single term might not strongly determine the relevance of a citation of important pair. Bigram terms have been proven more helpful in citation classification systems [14]; therefore, in this study, we form a list of bigram terms extracted from citation context (of cited paper) in citing paper. This is conducted based on the assumption that two consecutive terms can clearly depict their associated class. First of all, the bigram terms appearing in important pairs were extracted using the NLP library in python. The next step involves preparing a list for all the bigram terms labeled as “important” terms by a domain expert who is an Associate Professor in the field of Computer Science. The terms not labeled as important were excluded from the list. The list was developed from the citation context of 70% of pairs used for training (from both the data sets) which was then tested on the remaining 30% pairs using the algorithm below. In this case, any of the terms from the lists are matched with the bigram term appearing in a test pair, is assigned a weight of “1”. Similarly, the value of 1 adds up for each matched bigram term. In simpler words, for a given score type, e.g., bigrams, and a given ML classifier, scores for each citing/cited pair in the training set, and expert-based binary classification, are provided. The classifier trains on this and then predicts classification for the 30% of citing papers that were held back. The quality is then assessed by comparing to the domain expert’s classifications for that 30%. Algorithm 1 computes the bigram score of the pair. It takes a testing pair as an input, computes its bigram terms score, and returns it. The returned value is then kept as the bigram score of the input pair, which is then given to the classifiers for binary classification.

Algorithm 1: Bigram Score Computation of Paper-Citation Pairs

Input: P_{test} // A testing pair
Output: $BT_{\text{score}}(P)$
Extract Bigram terms from P
Initialization:
 $BT_{\text{train}} = \{BT_0, BT_1, BT_2, \dots, BT_m\}$ // bigram terms list generated by domain experts
 $BT(P_{\text{test}}) = \{T_0, T_1, T_2, \dots, T_q\}$ // bigram terms of testing pair
 $BT_{\text{score}}(P) = 0$ // bigram term score of pair P
 Loop $i = 0$ to n // iterate n testing pairs
 Loop $j = 0$ to m // iterate m bigram terms (BT) in annoated BT list
 Loop $k = 0$ to s // iterate s BT in testing pair
 {
 if $((P_{\text{test}}(k)) == BT_{\text{train}}(j))$ // matches the bigram term of testing pair i with the annotated bigram term
 $BT_{\text{score}}(P_i) = BT_{\text{score}}(P_i) + 1$
 }
 End Loop
 End Loop
 End Loop
 return $BT_{\text{score}}(P)$

3.2.3. PoS Score

Part-of-speech (PoS) tagging is performed to tag each word in a text into its corresponding PoS. To the best of our knowledge, none of the contemporary binary citation classification-based studies have assessed the potential of PoS in determining important citations. This study exploits PoS, including noun, verb, adjective, and adverb, appearing in a citation sentence. We believe that the mentioned PoS are sufficient to determine the importance of a citation; therefore, we have discarded all other PoS such as pronouns, determiners, etc. The idea here is to pick 70% of the pairs and form a separate list for each PoS, i.e., noun, adjective, adverb, and verb and obtain the lists labeled as “important” from the same domain expert who labeled Bigram terms explained in the above section. For PoS tagging, Stanford CoreNLP (shown in Figure 3) is utilized. Next, the four PoS extracted from the testing pair were matched with the corresponding PoS list annotated by the domain expert. For instance, the verbs extracted from the citation context of the testing pair (i.e., from both citing and cited paper) were matched with the list of verbs, and the same process is performed for the other three PoS. The PoS found in the remaining 30% of pairs are matched to the list stored separately for each PoS. As previously conducted for bigram terms matching, the same methodology is adopted here to match the PoS in citation sentences. Algorithm 2 computes the PoS score of the pairs. It accepts a testing pair as an input, extracts the four PoS from the citation sentences using Stanford coreNLP, and prepares a separate list for each PoS, as mentioned in Algorithm 2. Next, similar lists are picked, i.e., a list of nouns from the testing pair and a list of annotated nouns are considered, and term by term matching is performed. On each matching, the value of 1 is added to the score of the respective PoS. The process continues until all the four lists from testing pairs are term-by-term matched with the respective annotated list. The algorithm returns the score for all four PoS, which is given to the classifiers for binary classification.

Algorithm 2: PoS Score Calculation of Paper-Citation Pairs**Input:** P_{test} // A testing pair**Output:** $n_{\text{score}}(P)$, $a_{\text{score}}(P)$, $v_{\text{score}}(P)$, $av_{\text{score}}(P)$ *Extract PoS from P using Stanford CoreNLP***Initialization:** $L_1 = \{n_1, n_2, n_3, \dots, n_m\}$ // list of annotated nouns $L_2 = \{a_1, a_2, a_3, \dots, a_m\}$ // list of annotated adjectives $L_3 = \{v_1, v_2, v_3, \dots, v_m\}$ // list of annotated verbs $L_4 = \{av_1, av_2, av_3, \dots, av_m\}$ // list of annotated adverbs $TL_1 = \{tn_1, tn_2, tn_3, \dots, tn_m\}$ // presents nouns appearing in citation context of testing pair $TL_2 = \{ta_1, ta_2, ta_3, \dots, ta_m\}$ // presents adjective appearing in citation context of testing pair $TL_3 = \{tv_1, tv_2, tv_3, \dots, tv_m\}$ // presents verbs appearing in citation context of testing pair $TL_4 = \{tav_1, tav_2, tav_3, \dots, tav_m\}$ // presents adverbs appearing in citation context of testing pair $n_{\text{score}}(P) = 0$ // noun $a_{\text{score}}(P) = 0$ // adjectives $v_{\text{score}}(P) = 0$ // verbs $av_{\text{score}}(P) = 0$ // adverbs $PoS_u(P) = 0$ // score of PoS_u **repeat:**Loop $i = 0$ to n // iterate n testing pairsif $(L_{(u)} == TL_{(u)})$ // $u = 1, 2, 3, 4$ which ensures that same list should be picked for each pair (i.e., list of annotated nouns should be matched with the nouns appearing in citation context of the testing pairLoop $j = 0$ to m // iterate m BT in bigram terms listLoop $k = 0$ to s // iterate s BT in testing pair

{

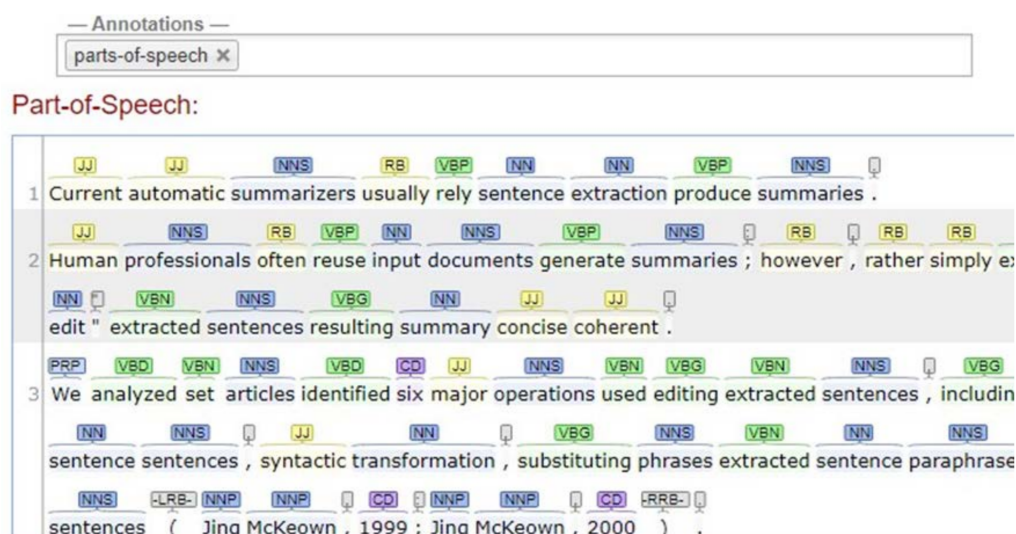
if $(P_{\text{test}}(L_{(u)}(k)) == TL_{(u)}(j))$ // matches the u PoS of testing pair i with u PoS list generated by domain experts $PoS_u(P) = PoS_u(P) + 1$

}

End Loop

End Loop

End Loop

return $PoS_u(P)$ **until** all four corresponding PoS lists are matched**Figure 3.** PoS tagging using Stanford CoreNLP.

3.2.4. Similarity Computation

In this module, the content similarity between paper-citation pairs is computed in two dimensions: (1) section-wise and (2) overall. The notion of this similarity computation is picked with an assumption that a high similarity count may determine important citation. There may be a probability that a high similarity score appears among certain logical sections of pairs, providing a solid clue regarding the citation class. Based on this assumption, we intend to scrutinize section-wise similarity behavior among pairs. In section-wise similarity computation, IMRaD sections of pairs are assessed based on their similarity score.

For instance,

- Introduction-Introduction (I-I)
- Methodology-Methodology (M-M)
- Results-Results (R-R)
- Discussion-Discussion (D-D)

The similarity is computed using above mentioned combinations of sections. The similarity is calculated using the cosine measure.

A. Cosine Similarity:

Cosine similarity is a metric that decides the similarity among two documents of variable sizes. The cosine similarity follows the notion that closer the documents by angle will lead to the high cosine value, which lies between 0 and 1. Equation (5) computes the cosine similarity between two documents.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (5)$$

where $A \cdot B = \sum_{i=1}^n A_i B_i = A_1 B_1 + A_2 B_2 + \dots + A_n B_n$ is the dot product among two vectors.

In the proposed study, “A” represents the content of the citing paper, and “B” is the cited paper’s content. It is pertinent to mention here that cosine similarity is computed in five ways: (1) cosine similarity between full content of citing and cited paper (2) Cosine similarity between “Introduction” sections of both citing and cited paper, (3) cosine similarity between “Methodology” sections of citing and cited paper, (4) cosine similarity between “Results” sections of citing and cited paper and (5) cosine similarity between “Discussion” sections of citing and cited paper.

3.2.5. Section-Wise Part of Speech (PoS)

Typically, a research paper encompasses different sections, often referred to as IMRaD. In this study, we intend to find the high existence of a particular PoS in the sections mentioned above. This experiment analyzes whether a specific PoS in a certain behavior helps determine the relationship between the pair. For this experiment, we consider four PoS, noun, verb, adverb, and adjective. The content of a section is pre-processed before the PoS extraction. In a pre-processing step, all the stop words are removed from the text using Onix Text Retrieval Toolkit (<https://rdrr.io/cran/qdapDictionaries/man/OnixTxtRetToolkitSWL1.html> (accessed on June 2021)). After that, PoS extraction is performed using Stanford CoreNLP. Figure 2 shows an example of a paper from Valenzuela’s data set on how Stanford CoreNLP labels terms into their corresponding PoS.

The objective of performing this experiment is to discover the patterns of PoS existence in important and incidental papers. It should be noted that all types of the four PoS, labeled by Stanford NLP, have been merged into a single PoS. For instance, all types of nouns such as proper nouns, abstract nouns, etc., have been combined into the category “noun”. The same has been performed for all the remaining PoS, i.e., verb, adjective, and adverb. It is pertinent to mention that authors do not strictly follow the same terminology for logical sections of research papers. For instance, some used “related work”, while others

used “Literature review” to describe state-of-the-art studies. In this experiment, the papers containing different terminologies for section names were labeled to the particular section of IMRaD by reading the section’s content. It should be noted that the purpose of utilizing this feature is to analyze the difference of appearing PoS in the same sections of important pairs and non-important pairs.

Following are notations explanations of the equations used to calculate the PoS score between four logical sections of citing papers. I represents “Introduction”, M represents “Methods”, R represents “Results”, D represents “Discussion,” and P denotes paper-citation “pair”.

Let $\langle S_{ij}, C_{ij} \rangle$ be a pair wherein S_i represents i th source (cited) paper and C_{ij} denotes the j th citing paper of S_i . Thus, S_{ij} represents i th source paper paired with j th citing the paper and, $i = \{1, 2, 3, \dots, n\}$ denotes the count of the source paper from 1 to n and $j = \{1, 2, 3, \dots, m\}$ the number of citing a paper from 1 to m . Thus, as explained earlier, this study determines the section-wise role of the four PoS in important and non-important pairs.

Equations (6)–(9) computes the appearance of “noun”, “verb”, “adjective” and “adverb” respectively, in “Introduction” sections of the cited paper (i.e., S_{ij}) and its citing paper C_{ij} .

$$I_{noun}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{noun}(S_{ij}) + PoS_{noun}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (6)$$

$$I_{verb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{verb}(S_{ij}) + PoS_{verb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (7)$$

$$I_{adjective}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adjective}(S_{ij}) + PoS_{adjective}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (8)$$

$$I_{adverb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adverb}(S_{ij}) + PoS_{adverb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (9)$$

Similarly, Equations (10)–(13) examines the existence of “noun”, “verb”, “adjective” and “adverb” respectively, in “Methods” sections of the cited paper (i.e., S_{ij}) and its citing paper C_{ij} .

$$M_{noun}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{noun}(S_{ij}) + PoS_{noun}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (10)$$

$$M_{verb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{verb}(S_{ij}) + PoS_{verb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (11)$$

$$M_{adjective}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adjective}(S_{ij}) + PoS_{adjective}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (12)$$

$$M_{adverb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adverb}(S_{ij}) + PoS_{adverb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (13)$$

The following Equations (14)–(17) are used to compute the scores of noun, verb, adjective, and adverb, respectively, in the “Results” sections of the cited paper (i.e., S_{ij}) and its citing paper C_{ij} .

$$R_{noun}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{noun}(S_{ij}) + PoS_{noun}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (14)$$

$$R_{verb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{verb}(S_{ij}) + PoS_{verb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (15)$$

$$R_{adjective}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adjective}(S_{ij}) + PoS_{adjective}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (16)$$

$$R_{adverb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adverb}(S_{ij}) + PoS_{adverb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (17)$$

Likewise, Equations (18)–(21) calculates the noun, verb, adjective, and adverb scores, respectively, in the “Discussion” sections of the cited paper (i.e., S_{ij}) and its citing paper C_{ij} .

$$D_{noun}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{noun}(S_{ij}) + PoS_{noun}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (18)$$

$$D_{verb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{verb}(S_{ij}) + PoS_{verb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (19)$$

$$D_{adjective}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adjective}(S_{ij}) + PoS_{adjective}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (20)$$

$$D_{adverb}(P_{ij}) = \left| \frac{\sum_{j=1}^m \{PoS_{adverb}(S_{ij}) + PoS_{adverb}(C_{ij})\}}{\sum_{j=1}^m \{PoS(S_{ij}) + PoS(C_{ij})\}} \right| \quad (21)$$

4. Results and Discussion

This section delineates results achieved by applying the proposed methodology along with their detailed analysis. Some of the research papers from the dataset by Valenzuela et al. [16] were not found on Association for Computational Linguistics (ACL) anthology; therefore, they have been discarded from the data set. The availability of both the data sets is similar, as stated in [14].

4.1. Classification

Once the above-listed features have been calculated by applying the proposed methodology, they are assigned as features’ scores to the machine learning tool Waikato Environment for Knowledge Analysis (WEKA) for classification. We have employed SVM, RF, and KLR classifiers, with 10-fold cross-validation using the WEKA tool. The configurations details of these classifiers are as follows: (1) Radial basis function (RBF) kernel with degree 2 for SVM, (2) 10 number of trees for RF and 0 maximum depth, and (3) KLR with degree 2. The classification outcomes are evaluated using standard evaluation parameters that contain recall, precision, and F-measure. The reason for choosing these measures is that the contemporary studies have evaluated their results using the measures, so it would be feasible to draw comparisons. The evaluation measures are represented in macro-average; therefore, the F-measure is not necessary to be relied upon between precision and recall.

4.2. Features’ Individual Performance

Firstly, we have scrutinized the individual potential of each feature and the best-performed classifier. Figure 4 shows the precision, recall, and F-measure values achieved against all the employed features. It is pertinent to mention that since our focus is to find the best performing binary classifier among the applied ones, we have only reported values of those classifiers for which the highest value of precision, recall, and F-measure is attained.

The mentioned classifiers are the ones that have outperformed the other classifiers used in this study. It can be seen that the highest value of F-measure (i.e., 0.71) is achieved by the feature M_M (i.e., the content similarity between methodology sections of a pair) from the section-wise similarity category, followed by the PoS based feature noun with 0.63 F-measure. The lowest F-measure score is 0.42, observed for the feature I_I (i.e., the

content similarity between Introduction sections of a pair). Figure 3 illustrates the outcomes attained by using Valenzuela’s data set.

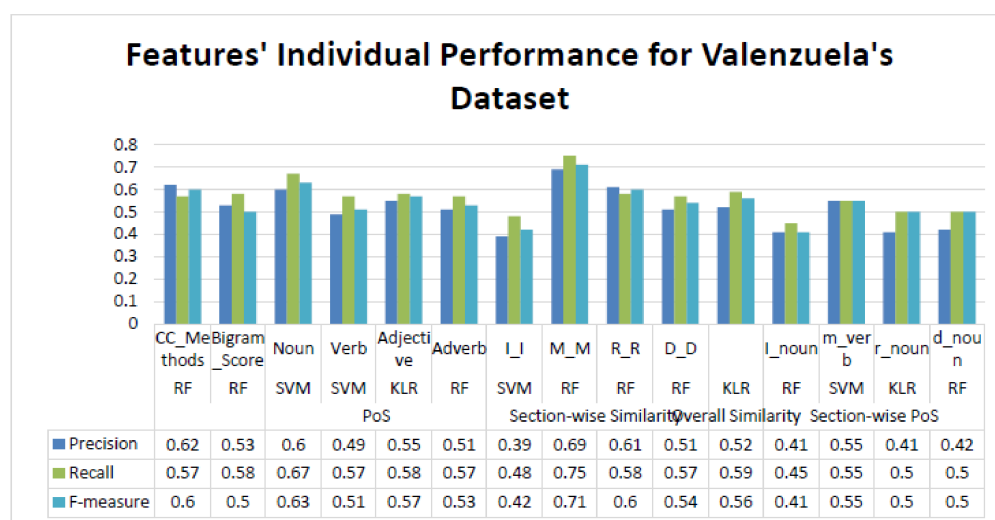


Figure 4. Individual performance of features for Valenzuela’s dataset.

Figure 5 shows details about precision, recall, and the F-measure score achieved by the harnessed features for Qayyum’s data set. The highest value of F-measure is 0.73, secured for CC_methodology by random forest classifier. The second top scored feature is Noun from the PoS category, having an F-measure of 0.71 for SVM, and the minimum F-measure score is achieved by adverb feature from PoS with an F-measure of 0.49.

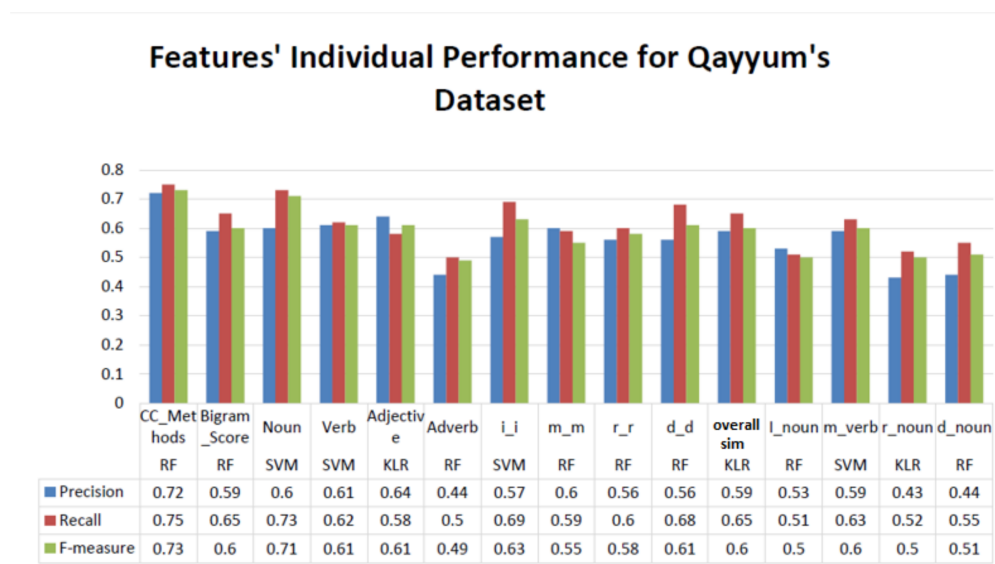


Figure 5. Individual performance of features for Qayyum’s dataset.

4.3. Features’ Combinations

To assess collective contributions of features towards binary citation classification, we have formed every possible combination of features ranging from double to combination of all features. The results achieved by combining all the features are reported in the comparison section. In this section, we have noted the best combination from all the remaining combinations. Figures 6 and 7 visualize the results of outperformed combinations for both the data sets. The combination of “Section-wise similarity (M_M) + CC_Methods” scored the highest with an F-measure of 0.73. For Qayyum’s data set, the best performance is

observed by the combination “CC_Methods + Noun” that attained an F-measure of 0.76. The top scored combinations for both the data sets were classified with an RF classifier. In both the data sets, the feature “CC_Methods” is present in the top-scored combination as shown in Equations (5) and (6), which indicates that considering the count of a citation in the methodology section has a strong influence in determining an important citation.

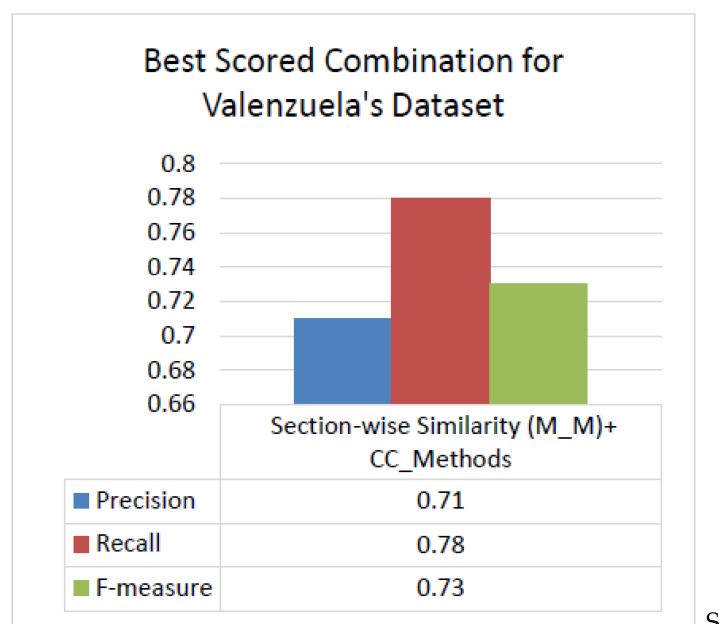


Figure 6. Top scored combinations for Valenzuela’s dataset.

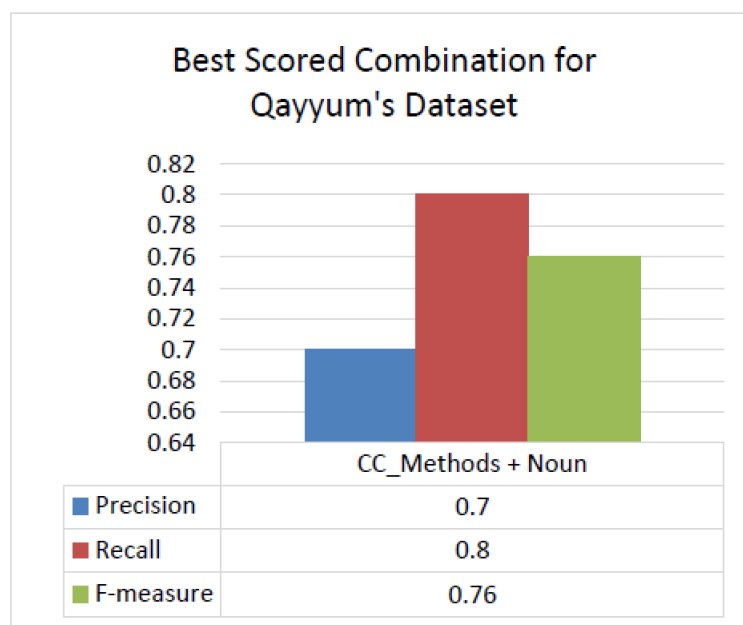


Figure 7. Top scored combinations for Qayyum’s dataset.

4.4. Comparisons

The results achieved by the proposed methodology are compared with three state-of-the-art techniques in binary citation classification [14,16].

The reasons for drawing a comparison with these three approaches are as follows:

- Valenzuela et al. [16]: In this study, we have harnessed the data set accumulated by Valenzuela et al. that has used content and metadata-based features.

- Qayyum et al. [14]: This study has also employed the same data sets and reported similar binary citation classification results using metadata-based features based on both the datasets employed in our proposed study.
- Nazir et al. [2]: Nazir et al. performed binary citation classification harnessing two same data sets used in our proposed study.

Since all of these studies have reported overall precision results, we have also drawn the comparisons using the precision score. The following table shows the precision scores achieved by combining all the features. Figure 8 shows that the proposed model achieves the highest value of precision as compared to existing studies for Valenzuela's data set. A precision of 0.88 is achieved by combining all the features in the proposed approach for Valenzuela's data set. This is the highest of all other precision scores.

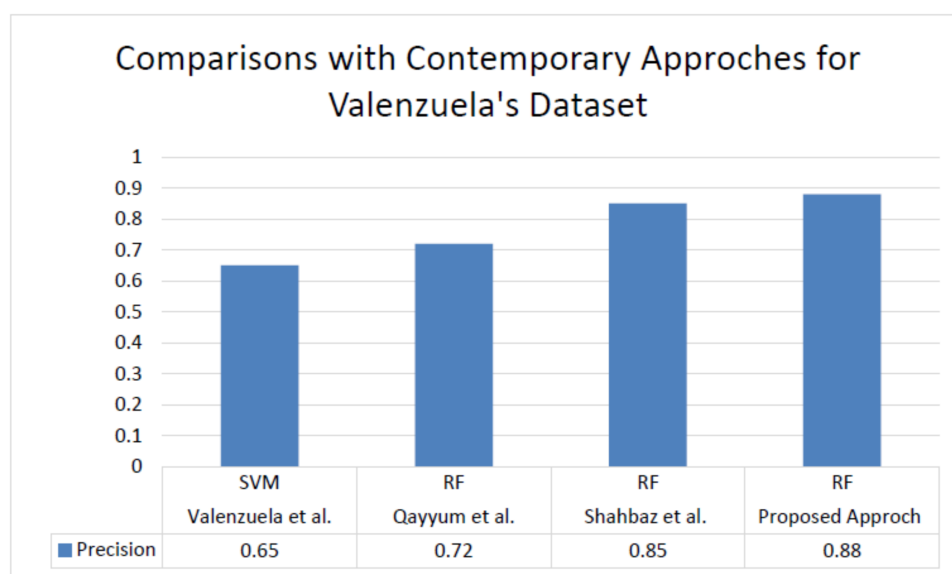


Figure 8. Comparison analysis with contemporary approaches for Valenzuela's dataset.

Similarly, the proposed methodology achieved the highest precision value for Qayyum's data set, as shown in Figure 9.

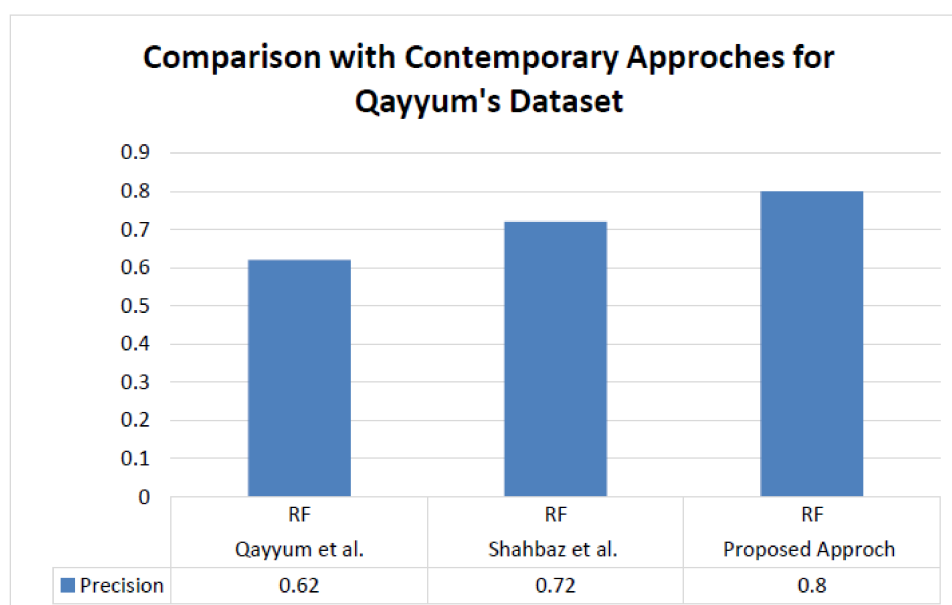


Figure 9. Comparison analysis with contemporary approaches for Qayyum's dataset.

Another essential aspect to be contemplated here is the result produced by the RF classifier remained consistent. The studies [2,14] have also reported that the RF classifier performed best in their proposed approaches.

The outcomes of the proposed study have revealed different insights to binary citation classification. Analysis of the performance varying from individual to collective contribution of the employed features shows significant potential in tackling important citations. From individual features, CC_method and similarity between M_M sections outperformed other features. These features were incorporated based on the assumption that citation count of cited paper in the methodology section of citing paper may depict an “important” relation between pairs as the methodology section usually contains comparatively a smaller number of citations and the ones that are part of this section are usually represent those papers which are very close to the citing paper. Similarly, the highest value of cosine similarity between the methodology section of pairs has also been proven quite helpful. This validates our assumption that cited and citing papers mostly use similar terms if they hold an “important” relation. Among these two features, CC_method could be considered more worthy as it is present in the top-scored combinations of features for both the employed data sets, as shown in Figure. Another important finding is the existence of the noun feature from the PoS group. To the best of our knowledge, no existing literature has the potential of PoS in finding important citations. The outcomes here suggest that the presence of ‘Noun’ in citation sentences of important pairs should be given considerable importance.

In this study, we manually formed a list of four PoS from 70% of the pairs because we were intended to find which of the four PoS has more presence in citing sentences of important pairs. The outcomes have ensured that “Noun” has more existence than the other three PoS, i.e., verb, adjective, adverb. In the future, only a high count of Noun in citation sentences could be deemed as a clue in determining important citations. Based on the highest precision value from existing studies, we claim that the identified list of features and proposed methodology holds strong potential for finding important citations.

5. Conclusions

There has been a continuous debate in the scientific community regarding filtering un-important reasons to refine the approaches wherein a mere count of citations is deemed a quintessential measure. Based on this argument, researchers have classified the citations into different reasons. Recently, the primary citation reasons have been converted into a small number of citation classes to identify only meaningful citations. Most of the schemes have preferred to exploit content-based features due to their diversity and richness. However, to the best of our knowledge, none of the existing studies produce sufficient accuracy. This paper has presented a comprehensive list of content-based features identified by critically analyzing the current state-of-the-art. The content of paper-citation pairs has been exploited to extract the required features, and then the proposed methodology is applied to classify citations into important and incidental classes. The classification has been performed using SVM, RF, and KLR. The outcomes yielded a precision of 0.80 and 0.88 for two different data sets. We claim that the proposed methodology has significant potential to tackle important citations.

Author Contributions: Conceptualization, F.Q. and H.J.; methodology, F.Q.; validation, F.Q., H.J., F.J. and D.-H.K.; formal analysis, F.Q.; investigation, F.Q.; resources, F.Q.; data curation, F.Q. and H.J.; writing—original draft preparation, F.Q. and F.J.; writing—review and editing, F.Q. and F.J.; visualization, F.Q. and H.J.; supervision, D.-H.K.; project administration, D.-H.K.; funding acquisition, D.-H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research was supported by Energy Cloud R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (2019M3F2A1073387), and this research was supported by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2018-0-01456, AutoMaTa: Autonomous Management framework based on artificial intelligent Technology for adaptive and disposable IoT). Any correspondence related to this paper should be addressed to Dohyeun Kim.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ziman, J.M. *Public Knowledge: An Essay Concerning the Social Dimension of Science*; Cambridge University Press: Cambridge, UK, 1968; Volume 519.
2. Nazir, S.; Asif, M.; Ahmad, S.; Bukhari, F.; Afzal, M.T.; Aljuaid, H. Important citation identification by exploiting content and section-wise in-text citation count. *PLoS ONE* **2020**, *15*, e0228885. [[CrossRef](#)] [[PubMed](#)]
3. Inhaber, H.; Przednowek, K. Quality of Research and the Nobel Prizes. *Soc. Stud. Sci.* **1976**, *6*, 33–50. [[CrossRef](#)]
4. Smith, A.T.; Eysenck, M. *The Correlation between RAE Ratings and Citation Counts in Psychology*; University of Royal Holloway: London, UK, 2002.
5. Hirsch, J.E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16569–16572. [[CrossRef](#)] [[PubMed](#)]
6. Ayaz, S.; Afzal, M. Identification of conversion factor for completing-h index for the field of mathematics. *Science* **2016**, *109*, 1511–1524. [[CrossRef](#)]
7. Ghani, R.; Qayyum, F.; Afzal, M.T.; Maurer, H. Comprehensive evaluation of h-index and its extensions in the domain of mathematics. *Science* **2019**, *118*, 809–822. [[CrossRef](#)]
8. Hashmi, A.M.; Qayyum, F.; Afzal, M.T. Insights to the state-of-the-art PDF Extraction Techniques. *IPSI Trans. Internet Res.* **2020**, *16*, 8.
9. Mazlounian, A.; Helbing, D.; Lozano, S.; Light, R.P.; Börner, K. Global multi-level analysis of the 'Scientific Food Web'. *Sci. Rep.* **2013**, *3*, 1167. [[CrossRef](#)]
10. Jamil, F.; Qayyum, F.; Alhelaly, S.; Javed, F.; Muthanna, A. Intelligent Microservice Based Blockchain for Healthcare Applications. *CMC Comput. Mater. Contin.* **2021**, *69*, 2513–2530.
11. Ali, A.; Iqbal, M.; Jamil, H.; Qayyum, F.; Jabbar, S.; Cheikhrouhou, O.; Baz, M.; Jamil, F. An Efficient Dynamic-Decision Based Task Scheduler for Task Offloading Optimization and Energy Management in Mobile Cloud Computing. *Sensors* **2021**, *21*, 4527. [[CrossRef](#)]
12. Jamil, F.; Kim, D. An Ensemble of a Prediction and Learning Mechanism for Improving Accuracy of Anomaly Detection in Network Intrusion Environments. *Sustainability* **2021**, *13*, 10057.
13. Jamil, F.; Kahng, H.K.; Kim, S.; Kim, D.H. Towards Secure Fitness Framework Based on IoT-Enabled Blockchain Network Integrated with Machine Learning Algorithms. *Sensors* **2021**, *21*, 1640. [[CrossRef](#)]
14. Qayyum, F.; Afzal, M.T. Identification of important citations by exploiting research articles' metadata and cue-terms from content. *Science* **2019**, *118*, 21–43. [[CrossRef](#)]
15. Garfield, E. Can citation indexing be automated. In *Statistical Association Methods for Mechanized Documentation, Symposium Proceedings*; NBS Miscellaneous Publications: Minneapolis, MN, USA, 1965; Volume 269, pp. 189–192.
16. Valenzuela, M.; Ha, V.; Etzioni, O. Identifying meaningful citations. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*; AAAI Publications: Palo Alto, CA, USA, 2015.
17. Zhu, X.; Turney, P.; Lemire, D.; Vellino, A. Measuring academic influence: Not all citations are equal. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 408–427. [[CrossRef](#)]
18. Teufel, S.; Siddharthan, A.; Tidhar, D. Automatic Classification of Citation Function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing—EMNLP '06*, Sydney, NSW, Australia, 22–23 July 2006; pp. 103–110.
19. Brooks, T.A. Private acts and public objects: An investigation of citer motivations. *J. Am. Soc. Inf. Sci.* **1985**, *36*, 223–229. [[CrossRef](#)]
20. Case, D.O.; Higgins, G. How can we investigate citation behavior? A study of reasons for citing literature in communication. *J. Am. Soc. Inf. Sci.* **2000**, *51*, 635–645. [[CrossRef](#)]
21. Spiegel-Rusing, I. Science studies: Bibliometric and content analysis. *Soc. Stud. Sci.* **1977**, *7*, 97–113. [[CrossRef](#)]
22. Oppenheim, C.; Renn, S.P. Highly cited old papers and the reasons why they continue to be cited. *J. Am. Soc. Inf. Sci.* **1978**, *29*, 225–231. [[CrossRef](#)]
23. Finney, B. The reference characteristics of scientific texts. Master's Thesis, The City University of London, London, UK, 1979.
24. Garzone, M.; Mercer, R.E. Towards an Automated Citation Classifier. In *Proceedings of the Lecture Notes in Computer Science*, Cagliari, Italy, 21–23 June 2000; Gabler: Wiesbaden, Germany, 2000; pp. 337–346.
25. Abu-Jbara, A.; Radev, D. Coherent citation-based summarization of scientific papers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, OR, USA, 19–24 June 2011; pp. 500–509.
26. Jochim, C.; Schütze, H. Towards a generic and flexible citation classifier based on a faceted classification scheme. In *Proceedings of the Proceedings of COLING*, Mumbai, India, 8–15 December 2012; pp. 1343–1358.

-
27. Hassan, S.-U.; Imran, M.; Iqbal, S.; Aljohani, N.R.; Nawaz, R. Deep context of citations using machine-learning models in scholarly full-text articles. *Science* **2018**, *117*, 1645–1662. [[CrossRef](#)]
 28. Moravcsik, J.M.; Murugesan, P. Some results on the function and quality of citations. *Soc. Stud. Sci.* **1975**, *5*, 88–91. [[CrossRef](#)]
 29. Nanba, H.; Okumura, M. Towards Multi-Paper Summarization Using Reference Information. *J. Nat. Lang. Process.* **1999**, *6*, 43–62. [[CrossRef](#)]
 30. Pride, D.; Knoth, P. Incidental or Influential?—Challenges in Automatically Detecting Citation Importance Using Publication Full Texts. In Proceedings of the Lecture Notes in Computer Science, Beer-Sheva, Israel, 29–30 June 2017; Gabler: Wiesbaden, Germany, 2017; Volume 10450, pp. 572–578.
 31. Tandon, N.; Jain, A. Citation context sentiment analysis for structured summarization of research papers. In Proceedings of the 35th German Conference on Artificial Intelligence, Saarbrücken, Germany, 24–27 September 2012.
 32. Ahmed, I.; Afzal, M.T. A Systematic Approach to Map the Research Articles' Sections to. IMRAD. *IEEE Access* **2020**, *8*, 129359–129371. [[CrossRef](#)]
 33. Porter, M.F. An algorithm for suffix stripping. *Program* **1980**, *14*, 130–137. [[CrossRef](#)]