

Towards Establishing a Research Lineage via Identification of Significant Citations

Tirthankar Ghosal^{§*}, Piyush Tiwary^{†*}, Robert Patton[‡] and Christopher Stahl[‡]

[§]Institute of Formal and Applied Linguistics, Charles University, CZ

[†]Indian Institute of Science, India

[‡]Oak Ridge National Laboratories, US

[§]ghosal@ufal.mff.cuni.cz

[†]piyushtiwary@iisc.ac.in

[‡](pattonrm, stahlcg)@ornl.gov

October 26, 2021

Abstract

Finding the lineage of a research topic is crucial for understanding the prior state of the art and advancing scientific displacement. The deluge of scholarly articles makes it difficult to locate the most relevant previous work. It causes researchers to spend a considerable amount of time building up their literature list. Citations play a crucial role in discovering relevant literature. However, not all citations are created equal. The majority of the citations that a paper receives provide contextual and background information to the citing papers. In those cases, the cited paper is not central to the theme of citing papers. However, some papers build upon a given paper, further the research frontier. In those cases, the concerned cited paper plays a pivotal role in the citing paper. Hence, the nature of citation the former receives from the latter is *significant*. In this work, we discuss our investigations towards discovering *significant citations* of a given paper. We further show how we can leverage significant citations to build a research lineage via a *significant citation graph*. We demonstrate the efficacy of our idea with two real-life case studies. Our experiments yield promising results with respect to the current state-of-the-art in classifying *significant citations*, outperforming the earlier ones by a relative margin of 20 points in terms of precision. We hypothesize that such an automated system can facilitate relevant literature discovery and help identify knowledge flow for a particular category of papers.

Keywords— citation classification, citation significance detection, machine learning, research lineage, citation graph, academic influence

1 Introduction

Literature searches are crucial to discover relevant publications. The knowledge discovery that ensues forms the basis of understanding a research problem, finding the previously explored frontiers, identifying research gaps, which eventually leads to the development of new ideas. However, with the exponential growth of scientific literature (including published papers and pre-prints) (Ghosal, Sonam, Ekbal, Saha, & Bhattacharyya, 2019), it is almost impossible for a researcher to go through the entire body of the scholarly works even in a very narrow domain. Citations play an important role here in finding the relevant articles that further topical knowledge. However, not all citations are equally (Zhu, Turney, Lemire, & Vellino, 2015) effective in finding relevant research. A majority of the papers cite a work contextually (Pride & Knoth, 2017a) for providing additional background context. Such background contextual citations help in the broader understanding; however, they are not central to the citing paper's theme. Some papers use the ideas in a given paper, build upon those ideas, and displace the body of relevant research. Such papers are expected to acknowledge the prior work (via citing them) duly. However, the nature of citation, in this case, is different from that of contextual citations. These citations, which heavily rely on a

*equal contribution

given work or build upon that work, are *significant citations*. However, the current citation count metric puts equal weights on all the citations. Therefore, it is inadequate to identify the papers that have significantly cited a given work and may have taken the relevant research forward. Identifying such *significant citations* are hence crucial to the literature study.

It is not uncommon that authors sometimes fail to acknowledge relevant papers' role in stemming up their ideas (Rousseau, 2007; Van Noorden, 2017). As a result, researchers spend a lot of their time searching for the most relevant papers to their research topic, thereby locating the subsequent papers that carried forward a given scientific idea. It is usually desirable for a researcher to understand the story behind a prior work and trace the concept's emergence and gradual evolution through publications, thereby identifying the knowledge flow. Researchers ideally curate their literature base by identifying significant references to a given paper and then hierarchically locating meaningful prior work.

The idea of recognizing *significant citations* is also important to understand the true impact of given research or facility. To understand how pervasive particular research was in the community, it is essential to understand its influence beyond the direct citations it received. To this end, tracking *transitive* influence of research via identifying significant citations could be one possible solution.

In this work, we develop automatic approaches to trace the lineage of given research via transitively identifying the significant citations to a given article. The overall objective of our work is two-fold:

- Accelerate relevant literature discovery via establishing a research lineage
- Find the true influence of a given work and its pervasiveness in the community beyond citation counts

There are two aspects to the problem: *identifying the relevant prior work* and *identifying the follow-up works that stemmed or are influenced by the current work*. The first aspect would facilitate relevant prior literature discovery for a paper. In contrast, the second aspect would facilitate discovering knowledge flow in subsequent relevant papers. Obviously, our approach would not be a *one shoe fits for all*. Still, we believe it is effective to find investigations that build upon relevant priors, facilitate relevant literature discovery, and thereby steer towards identifying the pervasiveness of a given piece of research in the community. We base our work to classify citations as *contextual* or *significant* and trace the lineage of research in a citation graph via identifying significant edges.

The major contributions of the current work are:

1. We use a set of novel and rich features to classify citations as significant or contextual.
2. A graph-based approach to trace the lineage of a given research work leveraging on citation classification.

2 Research Lineage

The mechanism of citations in academia is not always transparent (Van Noorden & Singh Chawla, 2019; Viuu, 2016; West, Stenius, & Kettunen, 2017). Problems like coercive citations (Wilhite & Fong, 2012), anomalous citations (Bai, Xia, Lee, Zhang, & Ning, 2016), citation manipulation (Bartneck & Kokkermans, 2011), *rich gets richer effects* (Ronda-Pupo & Pham, 2018), discriminatory citation practices (Camacho-Miñano & Núñez-Nickel, 2009), etc. has infested the academic community. However, in spite of all these known issues, citation counts and *h*-indices still remain the measures of research impact and tools for academic incentives, though long-debated by many (Cerdá, Nieto, & Campos, 2009; Laloë & Mosseri, 2009). Usually, we measure the impact of a given paper by the direct citations it receives. However, a given research may have induced a transitive effect on other papers, which is not apparent with the current citation count measures. Figure 1 shows a sample citation network where A could be a paper or a research facility. We want to know how pervasive was the research or facility A in the community. At $d=1$ are the direct citations to A. We see article B cites A significantly, or B is inspired by A. Other citations to A are background. At citation depth $d=2$, we see that article C and article D significantly cites B (direct citation). We see that C also cites

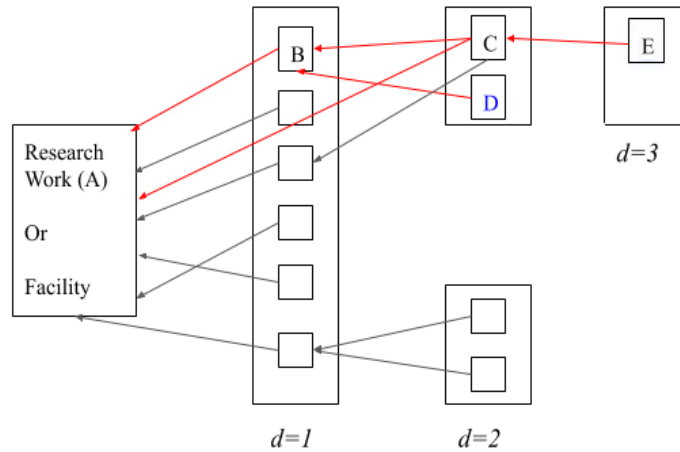


Figure 1: Research Lineage

A significantly. Finally, at citation depth $d=3$, E significantly cites C. We intend to understand if there is a lineage of research from A to E ($A \rightarrow B \rightarrow C \rightarrow E$). Although E does not cite A directly, can we identify A's influence on E? If E is a seminal work receiving hundreds of citations, can we infer that A was the prior work that indirectly inspired E? We are interested in discovering such hidden inspirations to honestly assess the contributions of a research article or a facility.

3 Related Work

Measuring academic influence has been a research topic since publications associate with academic prestige and incentives. Several metrics (Impact Factor, Eigen Factor, h -index, citation counts, alt metrics, etc.) came up to comprehend research impact efficiently. Still, each one is motivated on a different aspect and has found varied importance across disciplines. Zhu et al. (2015) did pioneering work on academic influence prediction leveraging on citation context. Shi, Wang, Chen, Liu, and Zhou (2019) presented a visual analysis of citation context-based article influence ranking. Xie, Sun, and Shen (2016) predicted paper influence in an academic network by taking into account the contents and venue of a paper, as well as the reputation of its authors. Shen et al. (2016) used topic modeling to measure academic influence in scientific literature. Manju, Kavitha, and Geetha (2017) identified influential researchers in an academic network using a rough-set based selection of time-weighted academic and social network features. Pileggi (2018) did a citation network analysis to measure academic influence. F. Zhang and Wu (2020) used a dynamic academic network to predict the future influence of papers. Ji, Tang, and Chen (2019) analyzed the impact of academic papers based on improved PageRank. F. Wang, Jia, Liu, and Liu (2019) assessed the academic influence of scientific literature via alt metrics. F. Zhao, Zhang, Lu, and Shai (2019) measured academic influence using heterogeneous author-citation networks. Recently, many deep learning-based methods are being explored for citation classification. Perier-Camby, Bertin, Atanassova, and Armetta (2019) attempt to compare deep learning-based methods with rule-based methods. They use deep learning-based feature extractors such as BCN (McCann, Bradbury, Xiong, & Socher, 2017) and ELMo (Peters et al., 2018) to extract semantic information and feed it to various classifiers for classification. They conclude that neural networks could be a potential dimension for citation classification when a large number of samples are available. However, for a small dataset like the one we use, rule-based methods clearly hold an advantage. Apart from this, the features used in rule-based methods are more comprehensible than features extracted from deep learning methods, thus providing deeper insights into analyzing factors that make a citation significant or contextual.

The closest literature for our task are the ones on citation classification. Citation classification has been explored in the works of Alvarez, Soriano, and Martínez-Barco (2017); Dong and Schäfer (2011); Qayyum and Afzal (2019); Teufel, Siddharthan, and Tidhar (2006). These works use features from the perspective of citation motivation. On the other hand there are works which emphasize on

features from semantic perspective. M. Wang et al. (2020) use syntactic and contextual information of citations for classification. Aljuaid, Iftikhar, Ahmad, Asif, and Afzal (2021); Amjad and Ihsan (n.d.) perform classification based on sentiment analysis of in-text citations. Athar (2011); Ihsan, Imran, Ahmed, and Qadir (2019) propose sentiment analysis of citations using linguistic studies of the citance. More recently, several open-source datasets for citation classification came up in the works of Cohan, Ammar, van Zuylen, and Cady (2019); Pride and Knoth (2020). Valenzuela, Ha, and Etzioni (2015) explored citation classification into *influential* and *incidental* using machine learning techniques which we adapt as *significant* and *contextual* respectively in this work.

In this work, we propose a rich set of features informed from both citation and context (semantics) perspectives, leveraging advantages of both types, thus performing better than all of the methods mentioned above. However, our problem is motivated beyond citation classification. We restrict our classification labels to *significant* and *contextual* unlike Valenzuela et al. (2015) as these labels are enough to trace the lineage of a work. Furthermore, to the best of our knowledge, we did not find any work leveraging citation classification for finding a research lineage. Hence, we only compare our performance for the citation significance detection sub-task with other approaches.

4 Dataset Description

We experiment with the Valenzuela dataset (Valenzuela et al., 2015) for our task. The dataset consists of incidental/influential human judgments on 630 citing-cited paper pairs for articles drawn from the 2013 ACL anthology, the full texts of which are publicly available. Two expert human annotators determined the judgment for each citation, and each citation was assigned a label. Using the author’s binary classification, 396 citation pairs were ranked as incidental citations, and 69 (14.3%) were ranked as influential (important) citations. For demonstrating our research lineage idea, we explore knowledge flow on certain papers of *Document-Level Novelty Detection* (Ghosal, Salam, Tiwari, Ekbal, & Bhattacharyya, 2018) and the High Performance Computing (HPC) algorithm *MENNDL* (Young, Rose, Karnowski, Lim, & Patton, 2015). Actual authors of these two topics helped us with manual annotation of their paper’s lineage.

5 Methodology

To identify significant citations, we pursue a feature-engineering approach to curate several features from cited-citing paper pairs. The objective is to classify the citations received by a given paper into SIGNIFICANT and CONTEXTUAL. The original cited citing papers in the Valenzuela dataset are in PDF. We convert the PDFs to corresponding XMLs using GROBID (Lopez, 2009). We use GROBID to parse our PDFs into XMLs as well as manually correct a few inconsistent files so that there is no discrepancy.

1. **Citation frequency inside the body of citing paper (F1):** We measure the number of times the cited paper is referenced from within the citing paper’s body. The intuition is that if a paper is cited multiple times, the cited paper may be significant to the citing paper.
2. **Are the authors of citing & cited paper the same? (Boolean) (F2):** We check if the authors of the citing and cited paper are the same. This might be the case of self-citation or can also signal the extension of the work.
3. **Author overlap ratio (F3):** This measures number of common authors in citing and cited paper normalized to the total number of authors in citing paper. Intuition is similar to F2.
4. **Is the citation occurring in a table or figure captions? (Boolean) (F4):** The intuition is that most of the citations in tables & figures appear for comparison/significantly referencing existing work. Hence, the citing paper might be an extension of the cited article or may have compared it with earlier significant work.
5. **Is the citation occurring in groups? (Boolean) (F5):** We check if the citation is occurring along with other citations in a group. Intuition is that such citations generally appear in related works to highlight a background detail; hence, they might not be a significant citation.

6. **Number of citations to the cited paper normalized by the total number of citations made by the citing paper (F6):** This measures number of citations to the cited paper by the citing paper normalized by the total number of citation instances in the citing paper. This measures how frequently the cited paper is mentioned compared to other cited papers in the citing paper.
7. **Number of citations to the cited paper normalized by the total number of bibliography items in the citing paper (F7):** This measures number of citations to the cited paper normalized to the total number of bibliography items in the citing paper. Intuition is similar to F6.
8. ***tf-idf* similarity between abstracts of the cited and citing paper (F8):** We take cosine similarity between the *tf-idf* representations of the abstracts of cited and citing papers. Intuition is that if the similarity is higher, the citing paper may be inspired/extended from the cited paper.
9. ***tf-idf* similarity between titles of the cited and citing paper (F9):** We take cosine similarity between the *tf-idf* representations of the titles of cited and citing papers.
10. **Average *tf-idf* similarity between citance and abstract of the cited paper (F10):** We calculate the similarity of each citance with the abstract of the cited article and take the average of it. Citances are sentences containing the citations in the citing paper. Citances reveal the purpose of the cited paper in the citing paper. Abstracts contain the contribution/purpose statements of a given paper. Hence similarity with citances may construe that the cited paper may have been used significantly in the current paper.
11. **Maximum *tf-idf* similarity between citance and abstract of the cited paper (F11):** We take the maximum of similarity of the citances (there could be multiple citation instances of the same paper in a given paper) with the abstract of the cited paper.
12. **Average *tf-idf* similarity between citance and title of the cited paper (F12):** We calculate the similarity of each citance with the title of the cited paper and take an average of it.
13. **Maximum *tf-idf* similarity between citance and title of the cited paper (F13):** We take the maximum of similarity of the citances with the title of the cited paper.
14. **Average Length of the Citance (F14):** Average length of the citances (in words) for multiple citances. Intuition is that if the citing paper has spent many words on the cited article, it may have significantly cited the corresponding article.
15. **Maximum Length of the Citance (F15):** Maximum length of the citances (in words).
16. **No. of words between citances (F16):** We take the average of the number of words between each pair of consecutive citance of the cited paper. This is set to 0 in the case of a single citance.
17. **In how many different sections does the citation appear in the citing paper? (F17):** We take the number of different sections in which the citation to cited paper occurs and normalize it with the total number of sections present in the citing paper. Intuition is that if a citation occurs in most sections, it might be a significant citation.
18. **Number of common references in citing & cited paper normalized by the total number of references in citing article (F18):** We count the number of common bibliographic items present in the citing & cited paper and normalize it with total bibliographic items present in the citing paper.
19. **Number of common keywords between abstracts of the cited and citing paper extracted by YAKE (Campos et al., 2018) (F19):** We compare the number of common keywords between abstract of citing & cited paper extracted using YAKE. Our instinct is that more number of common keywords would denote more similarity between abstracts.
20. **Number of common keywords between titles of the cited and citing paper extracted by YAKE (F20):** We compare the number of common keywords between the title of citing & cited paper extracted using YAKE.

21. **Number of common keywords between the body of the cited and citing paper extracted by YAKE (F21):** We compare the number of common keyword between the body of citing & cited paper extracted using YAKE.
22. **Word Mover’s Distance (WMD) (Huang et al., 2016) between abstracts of the cited and citing paper (F22):** We measure the WMD between abstracts of citing & cited paper. The essence of this feature is to calculate semantic distance/similarity between abstracts of the two papers.
23. **WMD between titles of the cited and citing paper (F23):** We measure the WMD between title of citing & cited paper.
24. **WMD between the body of the cited and citing paper (F24):** We measure the WMD between the body of citing & cited paper.
25. **Average WMD between citance and abstract of the cited and citing paper (F25):** We take the average of WMDs between citance and abstract of the cited paper.
26. **Maximum WMD between citance and abstract of the cited and citing paper (F26):** We take the maximum of WMDs between citance and abstract of the cited paper.
27. **Average VADER (Gilbert & Hutto, 2014) Polarity Index - Positive (F27), Negative (F28), Neutral (F29), Compound (F30):** We measure VADER polarity index of all the citance of cited paper, and take their average for each sentiment (positive, negative, neutral & compound).
28. **Maximum VADER Polarity Index - Positive (F31), Negative (F32), Neutral (F33), Compound (F34) of Citances:** We measure VADER polarity index of all the citance of cited paper, and take maximum among them for each sentiment (positive, negative, neutral & compound). The intuition to use sentiment information is to understand how the citing paper cites the cited paper.
29. **Number of common venues in Bibliography of citing and cited paper (F35):** We count the number of common venues mentioned in the bibliography of citing & cited paper and normalize it with the number of unique venues in citing paper. Higher venue overlap would signify that the papers are in the same domain (Ghosal, Sonam, et al., 2019).
30. **Number of common Authors in Bibliography of citing and cited paper (F36):** We count the number of common authors mentioned in the bibliography of citing & cited paper and normalize it with the number of unique authors in citing paper (Ghosal, Sonam, et al., 2019).

As mentioned earlier, only 14.3% of total citations are labeled as significant, which poses a *Class Imbalance* problem. To address this issue, we use SMOTE (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) along with random under-sampling of majority (contextual citation) class. We first split the dataset into 60% training & 40% testing data. Then we under-sample the majority class by 50%, and then we over-sample the minority class by 40%, on the training partition of the dataset.

6 Evaluation

Our evaluation consists of two stages: first, we evaluate our approach on the citation significance task. Next, we try to see if we can identify the research lineage via tracing significant citations across the two research topics (*Document-level Novelty* and *MENNDL*). We ask the original authors to annotate the lineage and verify it with our automatic method. We train our model on the Valenzuela dataset and use that trained model to predict significant citations of *Document-level Novelty* and *MENNDL* papers, thereby try to visualize the research lineage across the citing papers. We curate a small citation graph to demonstrate our idea. Please note that our task in concern is Citation Significance Detection, which is different from Citation Classification in literature. Whereas Citation Classification focuses on identifying the citation’s intent, Citation Significance aims to identify the value associated with the citation. Obviously, the two tasks are related to each other, but the objectives are different.

6.1 Citation Significance Detection

The goal of this task is to identify whether a citation was SIGNIFICANT or CONTEXTUAL. We experiment with several classifiers for the binary classification task such as kNN ($k = 3$), Support Vector Machines (kernel = RBF), Decision Trees (max_depth = 10) and Random Forest (n_estimators = 15, max_depth = 10). We found Random Forest to be the best performing one with our feature set. Table 1 shows our current results against the earlier reported results on the Valenzuela dataset. We attain promising results compared to earlier approaches with a relative improvement of 20 points in precision. Since the dataset is small, none of the earlier approaches or we attempted a deep neural approach on this dataset. Like us, Qayyum and Afzal (2019) also used Random Forest as the classifier; however, they relied on meta data features rather than content-based features for their work. Their experiments tried to answer: to what extent can the similarities and dissimilarities between metadata parameters serve as useful indicators for important citation tracking? which metadata parameters or their combinations are helpful in achieving good results? We specifically work with paper full-text content-based features, hence our approach leverages richer information since it takes into consideration the full-text of the works, whereas Qayyum and Afzal (2019) is solely based on metadata which helps us to achieve better performance.

Table 2 shows classification results of the various classifiers we experimented with. Clearly, our features are highly inter-dependent (Section 5), and hence it explains the better performance of Random Forests.

Methods	Precision
Valenzuela et al. (2015)	0.65
Qayyum and Afzal (2019)	0.72
Nazir, Asif, and Ahmad (2020)	0.75
Nazir, Asif, Ahmad, Bukhari, et al. (2020)	0.85
Current Approach	0.92

Table 1: Results on Citation Significance Detection on Valenzuela dataset

Methods	Precision	Recall	F1-Score	Accuracy
kNN	0.80	0.87	0.83	0.81
SVM	0.79	0.67	0.73	0.81
Decision Tree	0.80	0.82	0.81	0.86
Random Forest	0.92	0.82	0.87	0.90

Table 2: Classification Result of various Classifiers for Citation Significance

Figure 2 shows the importance of the top 10 features ranked as per their *information gain*. However, our experimental dataset is small, our features co-related, and hence it seems that some features have marginal contributions. We deem that in a real-life bigger dataset, the feature significance would be more visible. Here, we can see that features like *distance between citances*, *the number of concerned citation normalized by the total number of citations*, *similarity between cited-citing abstracts*, *in-text citation frequency*, *the similarity between citance & cited abstract*, play an important role in the classification. The other features featuring in the top 10 are: *distance between citance*, *number of citations from citing to cited normalized by the total citations made by the citing paper*, *the similarity between cited-citing abstracts*, *in-text citation frequency*, *the average similarity between citance & cited abstract*, *number of citations from citing to cited normalized by the total references made by the citing paper*, *number of common YAKE keywords between the body of citing and cited paper*, *the average similarity between citance and title of cited paper*, *the max similarity between citance and abstract of cited paper*, *neutral sentiment polarity of citance*. We explain the possible reasons behind the performance of these features in the subsequent sections. The precision with only using the top 10 features is 0.73. Hence, other features play a significant role, as well. A complete list of features and the corresponding information gain

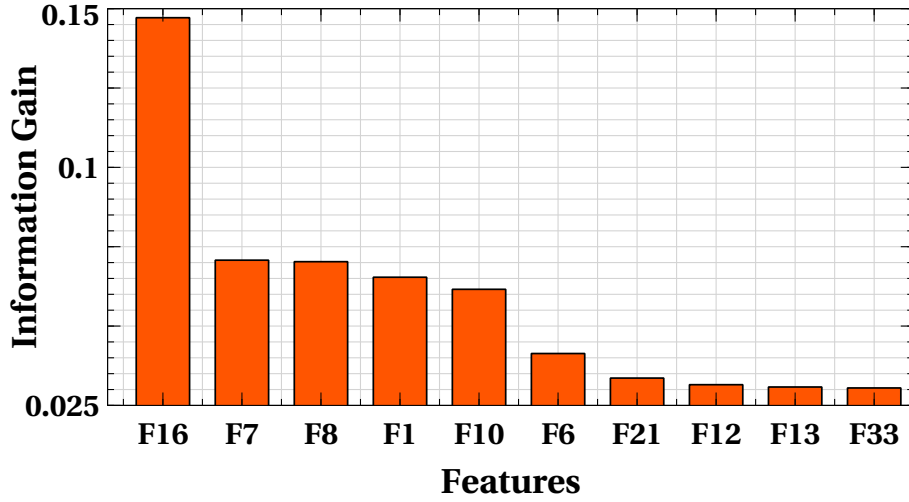


Figure 2: Feature importance ranked via Information Gain.

Feature	IG	Feature	IG	Feature	IG
F16	0.147	F24	0.024	F30	0.015
F7	0.070	F13	0.022	F27	0.015
F8	0.070	F33	0.021	F31	0.015
F1	0.065	F18	0.020	F32	0.014
F10	0.061	F3	0.020	F15	0.014
F6	0.041	F23	0.019	F9	0.013
F21	0.033	F35	0.019	F17	0.011
F12	0.031	F34	0.017	F36	0.011
F13	0.030	F19	0.017	F4	0.006
F33	0.030	F22	0.016	F5	0.006
F35	0.025	F28	0.016	F2	0.004
F26	0.024	F25	0.016	F20	0.003

Table 3: Information Gain (IG) due to each feature. Features are ranked in decreasing order of Information Gain.

is given in Table 3

To analyze the contribution of each feature, we evaluate our model using single feature at a time similar to Valenzuela et al. (2015). The precision after considering each feature individually is shown in Table 4. It is seen that the first 28 features in the table contribute significantly in classification, and the overall precision after considering all the features is even better (an improvement of 14 points). F1, F4 (suggesting that significant citations do occur in tables or figures) and F21 followed by F7, F19, and F3 are the best performing features. This indicates that features obtained from citation perspective are more useful. On the other hand the least performing features are F20 (perhaps due to small size of dataset), F13, F5 (suggesting significant citations also occur in groups), F17, F2, F22 and F33. Most of our observations are in line with Valenzuela et al. (2015).

To mention here, authors in Pride and Knoth (2017b) found *Number of Direct Citations*, *Author Overlap*, and *Abstract Similarity* to be the most important features. Our approach performs good enough to proceed with the next stage.

It is important to note that despite so many features, it is possible that some features might be correlated. Hence, we find the Pearson’s correlation coefficient between each pair of feature to see

Feature	Precision	Feature	Precision	Feature	Precision
F1	0.78	F15	0.28	F25	0.15
F4	0.76	F8	0.27	F11	0.14
F21	0.71	F10	0.27	F24	0.13
F7	0.68	F9	0.25	F31	0.11
F19	0.61	F27	0.23	F26	0.10
F3	0.50	F23	0.20	F33	0.08
F16	0.47	F36	0.20	F22	0.07
F28	0.43	F35	0.20	F2	0.04
F35	0.37	F12	0.19	F17	0.04
F6	0.33	F34	0.19	F5	0.03
F32	0.33	F30	0.17	F13	0.03
F33	0.29	F18	0.15	F20	0.01
Total					0.92

Table 4: Performance of Random Forest Model by using individual feature at a time. The features are listed in decreasing order of the precision.

how they are dependent on each other. The heatmap of correlation matrix is shown in Fig. 3.

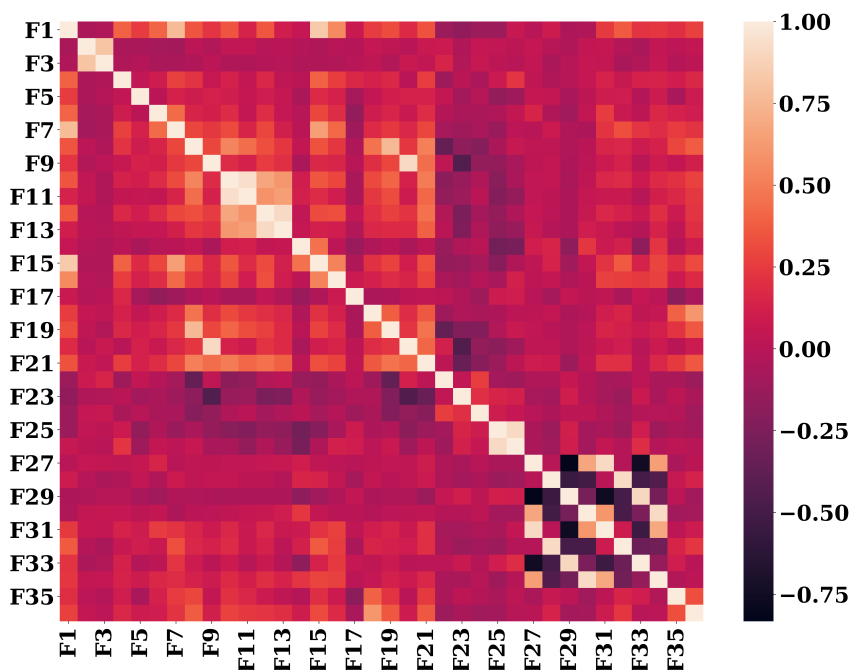


Figure 3: Heatmap of correlation between various pair of features.

We find that the average correlation coefficient between all the features is **0.074**. However, there are few pairs of features which have high correlation coefficients. We have listed such pairs in Table 5.

From Table 5 we can see that feature pairs like F10 & F11, F30 & F34, F28 & F32, F27 & F31, F12 & F13, F25 & F26, F29 & F33 have high correlation, which is understandable as these pairs are nothing but maximum and average of same quantity measured throughout the corresponding

Feature Pair	Correlation Coefficient	Feature Pair	Correlation Coefficient
F10 & F11	0.937	F25 & F26	0.910
F30 & F34	0.919	F9 & F20	0.907
F28 & F32	0.917	F29 & F33	0.905
F27 & F31	0.914	F1 & F15	0.835
F12 & F13	0.910	F27 & F29	0.832

Table 5: Feature pairs with high correlation coefficient.

literature. Hence, in order to reduce the complexity of the classifier, one may use just one of the features from each pair. The results after combining these features is shown in Table 6. It can be seen that even after clubbing these features there is no significant degradation in performance of our model.

Precision	Recall	F1 Score	Accuracy
0.91	0.80	0.85	0.89

Table 6: Citation significance results after combining features on the Valenzuela dataset

6.2 The 3C Dataset

As we mention earlier, the dataset used is small due to which the significance of each feature might not be visible explicitly. Hence, we also test our method on the 3C dataset which is relatively a bigger one. The 3C Citation Context Classification¹ Shared Task organized as part of the Second Workshop on Scholarly Document Processing @ NAACL 2021 is a classification challenge, where each citation context is categorized based on its purpose and influence. It consists of 2 subtasks:

- **Task A:** Multiclass classification of citation contexts based on purpose with categories - BACK-GROUND, USES, COMPARES CONTRASTS, MOTIVATION, EXTENSION, and FUTURE.
- **Task B:** Binary classification of citations into INCIDENTAL or INFLUENTIAL classes, i.e. a task for identifying the importance of a citation

The training and test dataset used for Task A and Task B are the same. The training data and test data consist of 3000 and 1000 instances, respectively. We use the data for Task B in for our experiments. However, the 3C dataset doesn't provide us with full text due to which we are only able to test only 19 of our features. We achieved an F1 score of **0.5358** with these 19 features on the privately-held 3C test set. Our relevant features in use here are: F1, F2, F9, F10, F11, F12, F13, F14, F15, F20, F23, F27, F28, F29, F30, F31, F32, F33, F34. We provide the results on the validation set using a random forest classifier in Table 7. The best performing system in 3C achieved an F1 score of 0.60 while the baseline F1 scores was 0.30.

Precision	Recall	F1 Score	Accuracy
0.569	0.575	0.572	0.606

Table 7: Citation Influence Classification Results on 3C Validation Set using a Random Forest Classifier

6.3 Research Lineage: Case Studies

Our end goal is not just citation classification but to make use of a highly accurate citation significance detection approach to trace significant citations and thereafter, try and establish a lineage of

¹<https://sdproc.org/2021/sharedtasks.html#3c>

the given research. As explained in Section 2, by research lineage we aim to identify the idea propagation via tracking the significant citations. To achieve this, we create a *Significant Citation Graph*. A Significant Citation Graph (SCG) is a graph-like structure, where each node represents a research paper. There is a directed edge between each cited-citing pair, whose direction is from cited paper node to citing paper node, indicating the flow of knowledge from cited paper to citing paper. In a usual case, all citations have equal weights in a citation graph. However, in our case, each edge is labeled as either *significant* or *contextual*, using the approach we discussed in the previous section. Our idea is similar to that of existing scholarly graph databases; however, we go one step further and depict how a particular concept or *knowledge* has propagated with consecutive citations.

Algorithm 1: Algorithm to Create Significance Citation Graph

Input: Trained Model & concerned research document, P
Output: Adjacency List for Citation Graph

```

1 Initialize adjacency list, A
2 Initialize an empty queue, Q
3 Q.add(P)
4 while Q is not empty do
5     for Each citation, C in Q[0] do
6         Extract features (F1-F36) for C
7         if C is Significant and C is not in Q then
8             Q.add(C)
9             A[Q[0]].add(C)
10    Q.pop()
11 return A

```

Algorithm 1 shows the method to create the adjacency list for the SCG. The Citation Significance Detection ML model is trained on a given dataset (Valenzuela in our case). To demonstrate the effectiveness of our method, we present a SCG for a set of papers on *Document-Level Novelty Detection* and *MENNDL*. Being the authors of the papers on these topics, we have identified the significant citations of each paper and used it to test the effectiveness of our proposed method to create an SCG.

6.3.1 Case Study I: Document-Level Novelty Detection

Figure 4 denotes an excerpt of a SCG from our *Document-Level Novelty Detection* papers. The red edges denote *significant* citations whereas black edges denote *contextual* citations. Our approach determined if a citation edge is *significant* or *contextual*. In the citation graph, we are interested in the lineage among four textual novelty detection papers (P1, P2, P3, P4), which the original authors annotate. We annotated that P1 is the *pivot paper* which introduced their document-level novelty detection dataset, and their other papers P2, P3, P4 are based on P1. While P2 and P4 address novelty classification, P3 aims to quantify textual novelty. Our approach conforms to the annotation by the original authors. With P1 as the pivot we can see that there are significant edges from P1 to each of P2, P3, and P4. There is also a significant edge between P2 and P4. However, there is no edge between P2 and P3, as they were contemporaneous submissions and their objective was different (P2 was about *novelty classification* and P3 was about *novelty scoring*). P1 → P2 → P4 forms a research lineage as P2 extends on P1 and P4 extends on P2. Furthermore, we see that P12, P25, P24, P22 (transitively) are some influential papers for P1. We verified from the authors that P25 was the paper to introduce the first *document-level novelty detection* dataset but from an information retrieval perspective. P25 inspired the authors to create the dataset in P1 for ML experiments. We construe that P12, P22, P24 had a significant influence on their investigations with P1. Hence, our approach (trained on a different set of papers in Valenzuela dataset) proved successful to identify the significant citations and thereby also identify the corresponding lineage.

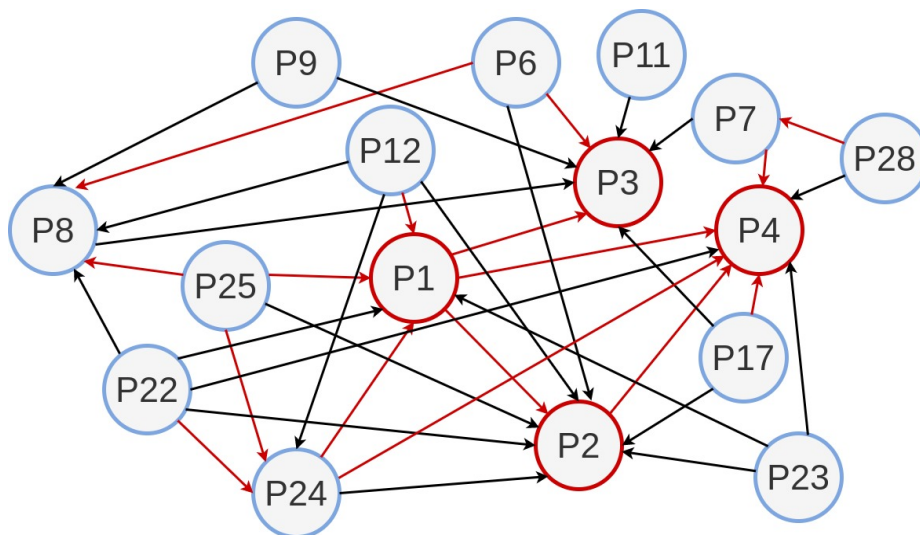


Figure 4: Significant Citation graph for a set of papers on *Document-Level Novelty Detection*. Please refer to the bibliography for the paper details. P1→Ghosal, Salam, et al. (2018), P2→Ghosal, Edithal, et al. (2018), P3→Ghosal, Shukla, et al. (2019), P4→Ghosal et al. (2020), P6→Soboroff and Harman (2003), P7→P. Zhao and Lee (2016), P8→Colomo-Palacios et al. (2010), P9→Tang et al. (2010), P11→Kusner et al. (2015), P12→Li and Croft (2005), P17→Schiffman and McKeown (2005), P22→Allan et al. (2003), P23→Soboroff and Harman (2005), P24→Karkali et al. (2013), P25→Y. Zhang et al. (2002), P28→F. Zhang et al. (2015)

6.3.2 Case Study II: MENNDL HPC Algorithm

We went ahead to test our approach’s efficacy to predict the *lineage* of a high-performance computing algorithm MENNDL. We show the research lineage of MENNDL Young et al. (2015) in Figure 5. We ask the original authors to annotate their research progression with MENNDL. As per the authors, the first paper to describe the MENNDL algorithm came in 2015, which is deemed as the *pivot* (P9). The follow-up paper that carried forward the work in P9 was P4 in 2017. Then P1 came in 2018 that built upon P4. P7, P12 came as extensions over P4. Next, P6 came in 2019 that took forward the work from P1. With P9 as the source, our approach correctly predicted the lineage as $P9 \rightarrow P4 \rightarrow P1 \rightarrow P6$. Also, the lineage $P9 \rightarrow P4 \rightarrow P12$ and $P9 \rightarrow P4 \rightarrow P7$ via tracing significant citations could be visible in the SCG at Figure 4. We annotate P8 as an application of P9; hence no significant link exists between P9 and P8.

From the above experiments and case studies, it is clear that our proposed method works reasonably well when a paper cites the influencing paper meaningfully. However, there are cases where some papers do not cite the papers from whom they are inspired. In such cases, our method would not work.

7 Conclusion and Future Work

Here, in this work, we present our novel idea towards finding a research lineage to accelerate literature review. We achieve *state-of-the-art* performance on citation significance detection, which is a crucial component to form the lineage. We leverage on that and show the efficacy of our approach on two completely different research topics. Our approach is simple and could be easily implemented on a large-scale citation graph (given the paper full-text). The training dataset is built from NLP papers. However, we demonstrate our approach’s efficacy by testing on two topics: one from NLP and the other from HPC, hence establishing that our approach is domain-agnostic. Identifying significant citations to form a research lineage would also help the community to understand the real impact of a research beyond simple citation counts. We would look forward to experimenting

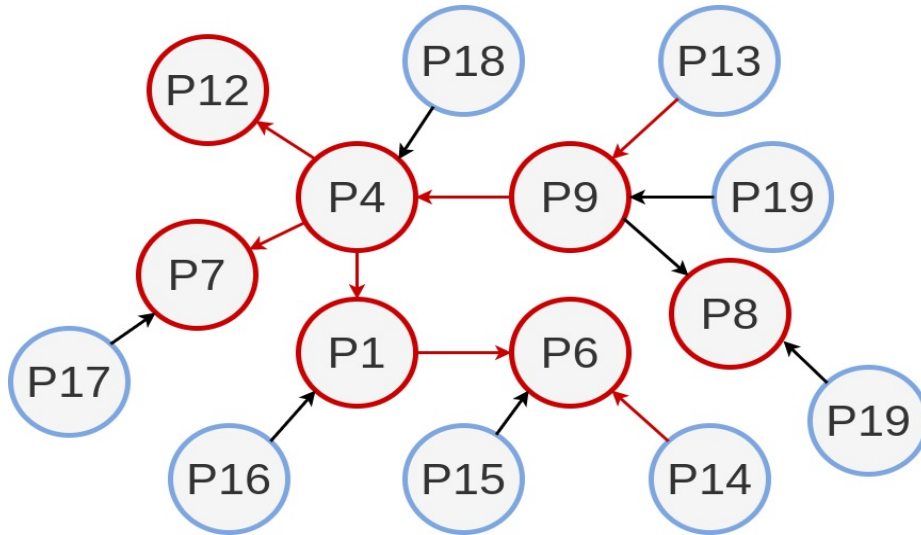


Figure 5: Significant Citation graph for a set of papers on *MENNDL HPC algorithm*. Please refer to the bibliography for the corresponding paper details. P1→Patton et al. (2018), P4→Young et al. (2017), P6→Patton et al. (2019), P7→J. T. Johnston et al. (2019), P8→T. Johnston et al. (2017), P9→Young et al. (2015), P12→Chae et al. (2019), P13→Jia et al. (2014), P14→Saltz et al. (2018), P15→Thorsson et al. (2018), P16→Bottou (2010), P17→Noh et al. (2015), P18→Lucchi et al. (2014), P19→Baldi et al. (2014), P20→Ciregan et al. (2012), P25→Y. Zhang et al. (2002), P28→F. Zhang et al. (2015)

with deep neural architectures to identify meaningful features for the current task automatically. Our next foray would be to identify the *missing citations* for papers which may have played instrumental role in certain papers but unfortunately are not cited. We release all the codes related to our experiment at <https://figshare.com/s/2388c54ba01d2df25f38>

8 Author’s Contribution

The first author conceptualized the work along with carrying out the investigation, formal analysis, data curation, methodology, baselines codes, and writing the original draft. The second author led the implementation, contributed in the formal analysis, and writing the paper draft. The third author was engaged in the overall supervision, funding acquisition, and providing resources for the investigation. The fourth author contributed towards the problem statement, data curation, writing-review and editing, and project administration.

9 Acknowledgement

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy (DOE). The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

The first author also thanks the Oak Ridge Institute for Science and Education (ORISE) to sponsor the first author for the Advanced Short-Term Research Opportunity (ASTRO) program at the Oak Ridge National Laboratory (ORNL). The ASTRO program is administered by the Oak Ridge

Institute for Science and Education (ORISE) for the U.S. Department of Energy. The first author also acknowledges the Visvesvaraya PhD fellowship award VISPHD-MEITY-2518 from Digital India Corporation under Ministry of Electronics and Information Technology, Government of India.

References

- Aljuaid, H., Iftikhar, R., Ahmad, S., Asif, M., & Afzal, M. T. (2021). Important citation identification using sentiment analysis of in-text citations. *Telematics and Informatics*, 56, 101492.
- Allan, J., Wade, C., & Bolivar, A. (2003). Retrieval and novelty detection at the sentence level. In *Proceedings of the 26th annual international acm sigir conference on research and development in information retrieval* (pp. 314–321).
- Alvarez, M. H., Soriano, J. M. G., & Martínez-Barco, P. (2017). Citation function, polarity and influence classification. *Nat. Lang. Eng.*, 23(4), 561–588. Retrieved from <https://doi.org/10.1017/S1351324916000346> doi: 10.1017/S1351324916000346
- Amjad, Z., & Ihsan, I. (n.d.). Verbnet based citation sentiment class assignment using machine learning.
- Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the acl 2011 student session* (pp. 81–87).
- Bai, X., Xia, F., Lee, I., Zhang, J., & Ning, Z. (2016). Identifying anomalous citations for objective evaluation of scholarly article impact. *PloS one*, 11(9), e0162364.
- Baldi, P., Sadowski, P., & Whiteson, D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature communications*, 5(1), 1–9.
- Bartneck, C., & Kokkermans, S. (2011). Detecting h-index manipulation through self-citation analysis. *Scientometrics*, 87(1), 85–98.
- Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of compstat'2010* (pp. 177–186). Springer.
- Camacho-Miñano, M., & Núñez-Nickel, M. (2009). The multilayered nature of reference selection. *J. Assoc. Inf. Sci. Technol.*, 60(4), 754–777. Retrieved from <https://doi.org/10.1002/asi.21018> doi: 10.1002/asi.21018
- Campos, R., Mangaravite, V., Pasquali, A., Jorge, A. M., Nunes, C., & Jatowt, A. (2018). Yake! collection-independent automatic keyword extractor. In *European conference on information retrieval* (pp. 806–810).
- Cerdá, J. H. C., Nieto, E. M., & Campos, M. L. (2009). What's wrong with citation counts? *D-Lib Magazine*, 15(3/4), 1082–9873.
- Chae, J., Schuman, C. D., Young, S. R., Johnston, J. T., Rose, D. C., Patton, R. M., & Potok, T. E. (2019). Visualization system for evolutionary neural networks for deep learning. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 4498–4502).
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357.
- Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3642–3649).
- Cohan, A., Ammar, W., van Zuylen, M., & Cady, F. (2019). Structural scaffolds for citation intent classification in scientific publications. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: Human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, volume 1 (long and short papers)* (pp. 3586–3596). Association for

- Computational Linguistics. Retrieved from <https://doi.org/10.18653/v1/n19-1361> doi: 10.18653/v1/n19-1361
- Colomo-Palacios, R., Tsai, F. S., & Chan, K. L. (2010). Redundancy and novelty mining in the business blogosphere. *The Learning Organization*.
- Dong, C., & Schäfer, U. (2011). Ensemble-style self-training on citation classification. In *Fifth international joint conference on natural language processing, IJCNLP 2011, Chiang Mai, Thailand, November 8-13, 2011* (pp. 623–631). The Association for Computer Linguistics. Retrieved from <https://www.aclweb.org/anthology/I11-1070/>
- Ghosal, T., Edithal, V., Ekbal, A., Bhattacharyya, P., Chivukula, S. S. S. K., & Tsatsaronis, G. (2020). Is your document novel? let attention guide you. an attention-based model for document-level novelty detection. *Natural Language Engineering*, 1–28. doi: 10.1017/S1351324920000194
- Ghosal, T., Edithal, V., Ekbal, A., Bhattacharyya, P., Tsatsaronis, G., & Chivukula, S. S. S. K. (2018). Novelty goes deep. a deep neural solution to document level novelty detection. In *Proceedings of the 27th international conference on computational linguistics* (pp. 2802–2813).
- Ghosal, T., Salam, A., Tiwari, S., Ekbal, A., & Bhattacharyya, P. (2018). Tap-dlnd 1.0: A corpus for document level novelty detection. *arXiv preprint arXiv:1802.06950*.
- Ghosal, T., Shukla, A., Ekbal, A., & Bhattacharyya, P. (2019). To comprehend the new: On measuring the freshness of a document. In *2019 international joint conference on neural networks (ijcnn)* (pp. 1–8).
- Ghosal, T., Sonam, R., Ekbal, A., Saha, S., & Bhattacharyya, P. (2019). Is the paper within scope? are you fishing in the right pond? In M. Bonn, D. Wu, J. S. Downie, & A. Martaus (Eds.), *19th ACM/IEEE joint conference on digital libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019* (pp. 237–240). IEEE. Retrieved from <https://doi.org/10.1109/JCDL.2019.00040> doi: 10.1109/JCDL.2019.00040
- Gilbert, C., & Hutto, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international conference on weblogs and social media (icwsm-14)*. available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf> (Vol. 81, p. 82).
- Huang, G., Guo, C., Kusner, M. J., Sun, Y., Sha, F., & Weinberger, K. Q. (2016). Supervised word mover’s distance. In *Advances in neural information processing systems* (pp. 4862–4870).
- Ihsan, I., Imran, S., Ahmed, O., & Qadir, M. A. (2019). A corpus-based study of reporting verbs in citation texts using natural language processing: Study of reporting verbs in citation texts using natural language processing. *CORPORUM: Journal of Corpus Linguistics*, 2(1), 25–36.
- Ji, C., Tang, Y., & Chen, G. (2019). Analyzing the influence of academic papers based on improved pagerank. In E. Popescu, T. Hao, T. Hsu, H. Xie, M. Temperini, & W. Chen (Eds.), *Emerging technologies for education - 4th international symposium, sete@icwl 2019, Magdeburg, Germany, September 23-25, 2019, revised selected papers* (Vol. 11984, pp. 214–225). Springer. Retrieved from https://doi.org/10.1007/978-3-030-38778-5_24 doi: 10.1007/978-3-030-38778-5_24
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd acm international conference on multimedia* (pp. 675–678).
- Johnston, J. T., Young, S. R., Schuman, C. D., Chae, J., March, D. D., Patton, R. M., & Potok, T. E. (2019). Fine-grained exploitation of mixed precision for faster cnn training. In *2019 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC)* (pp. 9–18).

- Johnston, T., Young, S. R., Hughes, D., Patton, R. M., & White, D. (2017). Optimizing convolutional neural networks for cloud detection. In *Proceedings of the machine learning on hpc environments* (pp. 1–9).
- Karkali, M., Rousseau, F., Ntoulas, A., & Vazirgiannis, M. (2013). Efficient online novelty detection in news streams. In *International conference on web information systems engineering* (pp. 57–71).
- Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015). From word embeddings to document distances. In *International conference on machine learning* (pp. 957–966).
- Laloë, F., & Mosseri, R. (2009). Bibliometric evaluation of individual researchers: not even right... not even wrong! *Europhysics News*, 40(5), 26–29.
- Li, X., & Croft, W. B. (2005). Novelty detection based on sentence level patterns. In *Proceedings of the 14th acm international conference on information and knowledge management* (pp. 744–751).
- Lopez, P. (2009). Grobid: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In *International conference on theory and practice of digital libraries* (pp. 473–474).
- Lucchi, A., Márquez-Neila, P., Becker, C., Li, Y., Smith, K., Knott, G., & Fua, P. (2014). Learning structured models for segmentation of 2-d and 3-d imagery. *IEEE transactions on medical imaging*, 34(5), 1096–1110.
- Manju, G., Kavitha, V., & Geetha, T. V. (2017). Influential researcher identification in academic network using rough set based selection of time-weighted academic and social network features. *Int. J. Intell. Inf. Technol.*, 13(1), 1–25. Retrieved from <https://doi.org/10.4018/IJIIIT.2017010101> doi: 10.4018/IJIIIT.2017010101
- McCann, B., Bradbury, J., Xiong, C., & Socher, R. (2017). Learned in translation: Contextualized word vectors. *arXiv preprint arXiv:1708.00107*.
- Nazir, S., Asif, M., & Ahmad, S. (2020). Important citation identification by exploiting the optimal in-text citation frequency. In *2020 international conference on engineering and emerging technologies (iceet)* (pp. 1–6).
- Nazir, S., Asif, M., Ahmad, S., Bukhari, F., Afzal, M. T., & Aljuaid, H. (2020). Important citation identification by exploiting content and section-wise in-text citation count. *PloS one*, 15(3), e0228885.
- Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1520–1528).
- Patton, R. M., Johnston, J. T., Young, S. R., Schuman, C. D., March, D. D., Potok, T. E., ... others (2018). 167-pflops deep learning for electron microscopy: from learning physics to atomic manipulation. In *Sc18: International conference for high performance computing, networking, storage and analysis* (pp. 638–648).
- Patton, R. M., Johnston, J. T., Young, S. R., Schuman, C. D., Potok, T. E., Rose, D. C., ... others (2019). Exascale deep learning to accelerate cancer research. In *2019 IEEE international conference on big data (big data)* (pp. 1488–1496).
- Perier-Camby, J., Bertin, M., Atanassova, I., & Armetta, F. (2019). A preliminary study to compare deep learning with rule-based approaches for citation classification. In *8th international workshop on bibliometric-enhanced information retrieval (bir) co-located with the 41st european conference on information retrieval (ecir 2019)* (Vol. 2345, pp. 125–131).
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- Pileggi, S. F. (2018). Looking deeper into academic citations through network analysis: popularity, influence and impact. *Univers. Access Inf. Soc.*, 17(3), 541–548. Retrieved from <https://doi.org/10.1007/s10209-017-0565-5> doi: 10.1007/s10209

-017-0565-5

- Pride, D., & Knoth, P. (2017a). Incidental or influential? - A decade of using text-mining for citation function classification. In J. Qiu, R. Rousseau, C. R. Sugimoto, & F. Xin (Eds.), *Proceedings of the 16th international conference on scientometrics and informetrics, ISSI 2017, wuhan, china, october 16-20, 2017* (pp. 1357–1367). ISSI Society.
- Pride, D., & Knoth, P. (2017b). Incidental or influential? - challenges in automatically detecting citation importance using publication full texts. In J. Kamps, G. Tsakonas, Y. Manolopoulos, L. S. Iliadis, & I. Karydis (Eds.), *Research and advanced technology for digital libraries - 21st international conference on theory and practice of digital libraries, TPDL 2017, thessaloniki, greece, september 18-21, 2017, proceedings* (Vol. 10450, pp. 572–578). Springer. Retrieved from https://doi.org/10.1007/978-3-319-67008-9_48 doi: 10.1007/978-3-319-67008-9_48
- Pride, D., & Knoth, P. (2020). An authoritative approach to citation classification. In R. Huang, D. Wu, G. Marchionini, D. He, S. J. Cunningham, & P. Hansen (Eds.), *JCDL '20: Proceedings of the ACM/IEEE joint conference on digital libraries in 2020, virtual event, china, august 1-5, 2020* (pp. 337–340). ACM. Retrieved from <https://doi.org/10.1145/3383583.3398617> doi: 10.1145/3383583.3398617
- Qayyum, F., & Afzal, M. T. (2019). Identification of important citations by exploiting research articles' metadata and cue-terms from content. *Scientometrics*, *118*(1), 21–43. Retrieved from <https://doi.org/10.1007/s11192-018-2961-x> doi: 10.1007/s11192-018-2961-x
- Ronda-Pupo, G. A., & Pham, T. (2018). The evolutions of the rich get richer and the fit get richer phenomena in scholarly networks: the case of the strategic management journal. *Scientometrics*, *116*(1), 363–383.
- Rousseau, R. (2007). The influence of missing publications on the hirsch index. *Journal of Informetrics*, *1*(1), 2–7.
- Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., ... others (2018). Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports*, *23*(1), 181–193.
- Schiffman, B., & McKeown, K. (2005). Context and learning in novelty detection. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 716–723).
- Shen, J., Song, Z., Li, S., Tan, Z., Mao, Y., Fu, L., ... Wang, X. (2016). Modeling topic-level academic influence in scientific literatures. In M. Khabsa, C. L. Giles, & A. D. Wade (Eds.), *Scholarly big data: AI perspectives, challenges, and ideas, papers from the 2016 AAAI workshop, phoenix, arizona, usa, february 13, 2016* (Vol. WS-16-13). AAAI Press. Retrieved from <http://www.aaai.org/ocs/index.php/WS/AAAIW16/paper/view/12598>
- Shi, C., Wang, H., Chen, B., Liu, Y., & Zhou, Z. (2019). Visual analysis of citation context-based article influence ranking. *IEEE Access*, *7*, 113853–113866. Retrieved from <https://doi.org/10.1109/ACCESS.2019.2932051> doi: 10.1109/ACCESS.2019.2932051
- Soboroff, I., & Harman, D. (2003). Overview of the trec 2003 novelty track. In *Trec* (pp. 38–53).
- Soboroff, I., & Harman, D. (2005). Novelty detection: the trec experience. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 105–112).
- Tang, W., Tsai, F. S., & Chen, L. (2010). Blended metrics for novel sentence mining. *Expert Systems with Applications*, *37*(7), 5172–5177.
- Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation func-

- tion. In D. Jurafsky & É. Gaussier (Eds.), *EMNLP 2006, proceedings of the 2006 conference on empirical methods in natural language processing, 22-23 july 2006, sydney, australia* (pp. 103–110). ACL. Retrieved from <https://www.aclweb.org/anthology/W06-1613/>
- Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Yang, T.-H. O., ... others (2018). The immune landscape of cancer. *Immunity*, 48(4), 812–830.
- Valenzuela, M., Ha, V., & Etzioni, O. (2015). Identifying meaningful citations. In C. Caragea et al. (Eds.), *Scholarly big data: AI perspectives, challenges, and ideas, papers from the 2015 AAAI workshop, austin, texas, usa, january, 2015* (Vol. WS-15-13). AAAI Press. Retrieved from <http://aaai.org/ocs/index.php/WS/AAAIW15/paper/view/10185>
- Van Noorden, R. (2017). The science that's never been cited. *Nature*, 552.
- Van Noorden, R., & Singh Chawla, D. (2019). Hundreds of extreme self-citing scientists revealed in new database. *Nature*, 572(7771), 578–579.
- Viiu, G.-A. (2016). A theoretical evaluation of hirsch-type bibliometric indicators confronted with extreme self-citation. *Journal of Informetrics*, 10(2), 552–566.
- Wang, F., Jia, C., Liu, J., & Liu, J. (2019). Dynamic assessment of the academic influence of scientific literature from the perspective of altmetrics. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *Proceedings of the 17th international conference on scientometrics and informetrics, ISSI 2019, rome, italy, september 2-5, 2019* (pp. 2528–2529). ISSI Society.
- Wang, M., Zhang, J., Jiao, S., Zhang, X., Zhu, N., & Chen, G. (2020). Important citation identification by exploiting the syntactic and contextual information of citations. *Scientometrics*, 125(3), 2109–2129.
- West, R., Stenius, K., & Kettunen, T. (2017). Use and abuse of citations. *Addiction Science: A Guide for the Perplexed*, 191.
- Wilhite, A. W., & Fong, E. A. (2012). Coercive citation in academic publishing. *Science*, 335(6068), 542–543.
- Xie, Y., Sun, Y., & Shen, L. (2016). Predicating paper influence in academic network. In *20th IEEE international conference on computer supported cooperative work in design, CSCWD 2016, nanchang, china, may 4-6, 2016* (pp. 539–544). IEEE. Retrieved from <https://doi.org/10.1109/CSCWD.2016.7566047> doi: 10.1109/CSCWD.2016.7566047
- Young, S. R., Rose, D. C., Johnston, T., Heller, W. T., Karnowski, T. P., Potok, T. E., ... Miller, J. (2017). Evolving deep networks using hpc. In *Proceedings of the machine learning on hpc environments* (pp. 1–7).
- Young, S. R., Rose, D. C., Karnowski, T. P., Lim, S.-H., & Patton, R. M. (2015). Optimizing deep learning hyper-parameters through an evolutionary algorithm. In *Proceedings of the workshop on machine learning in high-performance computing environments* (pp. 1–5).
- Zhang, F., & Wu, S. (2020). Predicting future influence of papers, researchers, and venues in a dynamic academic network. *J. Informetrics*, 14(2), 101035. Retrieved from <https://doi.org/10.1016/j.joi.2020.101035> doi: 10.1016/j.joi.2020.101035
- Zhang, F., Zheng, K., Yuan, N. J., Xie, X., Chen, E., & Zhou, X. (2015). A novelty-seeking based dining recommender system. In *Proceedings of the 24th international conference on world wide web* (pp. 1362–1372).
- Zhang, Y., Callan, J., & Minka, T. (2002). Novelty and redundancy detection in adaptive filtering. In *Proceedings of the 25th annual international acm sigir conference on research and development in information retrieval* (pp. 81–88).
- Zhao, F., Zhang, Y., Lu, J., & Shai, O. (2019). Measuring academic influence using heterogeneous author-citation networks. *Scientometrics*, 118(3), 1119–1140. Retrieved from

<https://doi.org/10.1007/s11192-019-03010-5> doi: 10.1007/s11192-019-03010-5

Zhao, P., & Lee, D. L. (2016). How much novelty is relevant? it depends on your curiosity. In *Proceedings of the 39th international acm sigir conference on research and development in information retrieval* (pp. 315–324).

Zhu, X., Turney, P. D., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *J. Assoc. Inf. Sci. Technol.*, 66(2), 408–427. Retrieved from <https://doi.org/10.1002/asi.23179> doi: 10.1002/asi.23179