Identifying Important Citations using Contextual Information from Full Text

Saeed-Ul Hassan
Department of Computer Science
Information Technology University
Ferozepur Road, Lahore 54600
Pakistan
saeed-ul-hassan@itu.edu.pk

Anam Akram
Department of Computer Science
Information Technology University
Ferozepur Road, Lahore 54600
Pakistan
anum.akram@itu.edu.pk

Peter Haddawy
Faculty of ICT
Mahidol University
Salaya, Nakhonpathom 73170
Thailand
peter.had@mahidol.ac.th

ABSTRACT

In this paper we address the problem of classifying cited work into important and non-important to the developments presented in a research publication. This task is vital for the algorithmic techniques that detect and follow emerging research topics and to qualitatively measure the impact of publications in increasingly growing scholarly big data. We consider cited work as important to a publication if that work is used or extended in some way. If a reference is cited as background work or for the purpose of comparing results, the cited work is considered to be non-important. By employing five classification techniques (Support Vector Machine, Naïve Bayes, Decision Tree, K-Nearest Neighbors and Random Forest) on an annotated dataset of 465 citations, we explore the effectiveness of eight previously published features and six novel features (including context based, cue words based and textual based). Within this set, our new features are among the best performing. Using the Random Forest classifier we achieve an overall classification accuracy of 0.91 AUC.

KEYWORDS

Machine learning, full text, citation context analysis, citation classification, mining scientific papers

1 INTRODUCTION

In this paper we address the problem of classifying cited work into important and non-important to the developments presented in a research publication. This task is vital for the algorithmic techniques that detect and follow emerging research topics and to qualitatively measure the impact of publications in increasingly growing scholarly big data.

When documenting new scientific discoveries, scholars cite earlier published work. Citation counts of a paper have been commonly used as a measure of the impact of the published work. But work may be cited for many reasons and all cited references do not necessarily influence the citing paper equally. Cited work that is being directly used or extended has more significant impact on a published piece of research than work that is cited to provide context, for comparison purposes, or as historical reference.

In this paper we classify cited work into important and non-important using 14 features: eight from previous work and six new features. We experiment with five different classification techniques: Support Vector Machine (SVM), Naïve Bayes, Decision Tree, K-Nearest Neighbors (KNN) and Random Forest. We use the dataset annotated by Valenzuela et al. [1] that labels 465 citations into two classes Important and Incidental (non-important). The dataset is publicly available and provides cited and citing paper links with the corresponding classes. Table 1 provides a description of the dataset.

Similar to the work of Valenzuela et al. [1] and Zhu et al. [2], we use features like citations from citing to cited paper, citations per section, and author overlap. To these we add additional features such as a comprehensive list of cue words for each class extracted from more than 80 research articles, and the similarity between the cited paper's abstract and the text around the citation in the body of citing paper. The research articles used to extract the cue words consist of papers discussed in the literature review section and selected papers found in their references. We evaluate our work following a comprehensive methodology. Our classifier has a predictive accuracy of 0.91 Area Under Receiver Operating Characteristic Curve (AUROC) and 0.84 Area Under the Precision-Recall Curve (AUCPR) using the Random Forest classifier.

Table 1: Statistics of Annotated Data

Dataset (label)		Annotated Citations		
		Initially four classes	Merged into two classes	No of citations (% of 465)
0	0	Related work	Incidental	398 (85.4)
	1	Comparison	Incidental	
1	2	Using the work	Important	67 (14.6)
	3	Extending the work	Important	

2 RELATED WORK

Researchers have recently proposed a number of citation classification schemes. Finney [3] proposed the following classification based on citation location in the document and the cue verbs around the citation context: 'assumed knowledge reference', 'confirmational reference', 'developmental reference', 'future research reference' and 'methodological reference', and 'tentative reference'. Garzone & Mercer [4] enhance this classification to 35 categories consisting of ten general types namely: affirmational type, negational type, tentative type, assumptive type, developmental type, methodological type, future research type, contrastive type, citations that utilize conceptual material and reader alert type. Oppenhe et al. [5] generate review articles by using models to classify research papers using citation links and citation types. Using a proposed 'words in common' method, they classify citations into two categories: Types S (supporting other studies) and Type P (part of other studies). They create a prototype system called PRESERI as a support system to write review articles.

Teufel et al. [6] propose a model to automatically classify citation functions into nine categories grouped as positive, negative and neutral. Their sample data contains 116 randomly selected articles and 2829 citation instances. Each citation instance is manually annotated to one of the categories and the papers are then automatically processed using machine-learning techniques. The results are then compared with the tags provided by the annotators. Dong & Schafer [7] introduce a citation classification scheme with four general categories: background, fundamental idea, technical basis and comparison. They use domain independent features from textual, physical and syntactic aspects to distinguish these citation functions. They compare their classification results with those of two human annotators. The whole corpus is divided into training and test datasets, with training ratios of 80%, 60%, 20% and 10%. The supervised model of Naïve Bayes is most effective as it is least sensitive to the size of the training set and performs well even for the training ratio of 10%. Moreover, an ensemble-style self-training algorithm is also developed to use and extend training data effectively.

Jochim & Schütze [8] introduce new features designed for citation classification and compare them experimentally with previously proposed citation features. They use MM [9] facet classification for citations. They show that their new features improve classification accuracy. The following features are used for citation classification: lexical, word-level linguistic, linguistic structure, location, frequency, sentiment, self-reference, and name entity recognition features. Tandon & Jain [10] propose a model that uses citation contexts from research articles to automatically generate a structured summary of a given article. A language model approach is used to categorize the citation contexts into five classes (summary, related work, extensions, limitations and strengths). The language models are constructed for each of the five classes and then the probability for a model to generate a specific citation context is estimated. Since each citation context can be categorized into more than one class, all models exhibiting the highest probability of generating a particular citation context are used for classification. Citation contexts are taken from the Microsoft Academic research engine and 500 such contexts are manually annotated to form the training set for the language models. Although an average precision of only 68.5% was achieved, the researchers state that increasing and refining the training set could improve the results.

Sula & Miller [11] introduce an experimental tool that extracts citation contexts from articles and categorizes them as positive or negative using a Naïve Bayes classifier based on two training sets (one positive, one negative). The corpus consists of 159 articles and 5700 citation contexts from four prominent humanities journals (Art Bulletin, Journal of Philosophy, Language and Modern Language Association of America). Zhu et al. [2] present a model to identify citations that have a central academic influence on the citing papers. The model utilizes supervised machine learning to predict academic influence using count-based, similarity-based, context-based and position-based features. The count-based features produced the best results. An influence-primed h-index is also proposed to measure the importance of a citation depending upon the number of times it had been mentioned in the body of the citing paper. Their experiments show that this indicator performs better than the conventional h-index.

Our work is closest to that of Valenzuela et al. [1], which focuses on identifying important citations in scholarly literature. They model this task as a supervised classification problem at two levels of detail: a coarse level which classifies citations as important and non-important, and a more detailed one which classifies citations into four categories: Incidental (related work), Incidental (comparison), Important (using the work) and Important (extending the work). Their classification mechanism utilizes a set of features that range from citation counts to where the citation appears in the body of the paper. They achieve 0.80 AUCPR on a dataset of 465 annotated citations. Our model outperforms their results by enhancing the feature set, including a comprehensive list of cue words for each class extracted from more than 80 research articles, and similarity between the cited paper's abstract and the text around reference in the body of citing paper. In addition - to the best of our knowledge - our employed evaluation experimental setup is most exhaustive that has ever deployed by existing related research.

3 DATA & METHOD

We use the publicly available annotated data from the work of Valenzuela et al. [1]. This data consists of 20,527 papers from the Association for Computational Linguistics anthology (http://allenai.org/data.html). In these papers, there are 106,509 total citations from which 465 are randomly annotated. Table 1 lists the four labeled classes and amounts of data in the classes. To form a binary class, label 0 (related work) and 1 (Comparison) are merged as 0 to indicate the Incidental class and label 2 (using the work) and 3 (extending the work) are merged as 1 to represent the Important class. In the dataset, 14.6% of the citations belong to the Important class and 85.4% to the Incidental class.

3.1 Approach

Our model revolves around three main aspects. First, a reference is important to a paper if it has been cited several times in the body of citing paper. Second, the position of a citation influences its important, e.g. a reference cited in the method section is likely to indicate that the current work is using or extending the cited work. In contrast, a citation in the related work section probably indicates an incidental citation. Third, the text around the citation in the body of citing paper is important since it can indicate the reason for the citation. Based on these aspects we select a set of features to classify the importance of cited work to the citing paper. Fig. 1 shows a flowchart of our deployed approach.

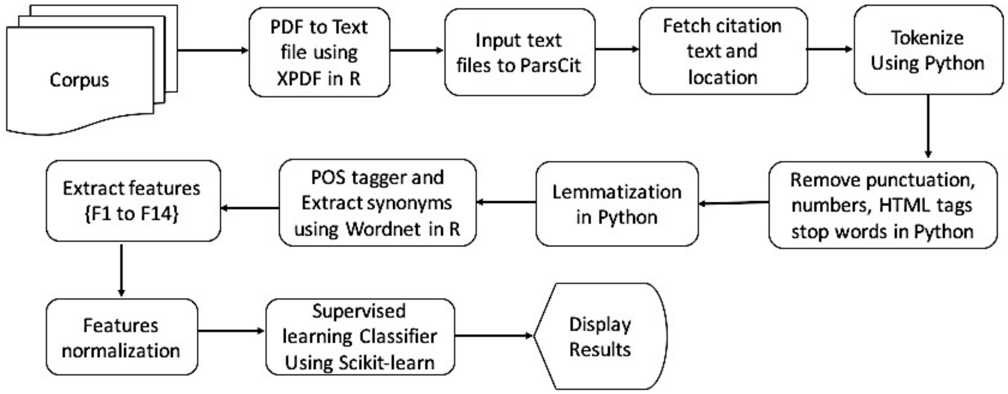

Figure 1: Flowchart of deployed approach

3.2 Features

We represent each cited work as a set of 14 features shown in Table 2, grouped as Context based, Cue word based and Textual. We use the following eight features, identified as most significant, by Valenzuela et al. [1]: F2 to F8, and F14.

Table 2: Features for Classification

Type		Features
Context Based Features	F1	Total number of citations received by a reference
	F2	Number of citations from the current paper to the cited paper
	F3	Citations in introduction section
	F4	Citations in literature review section
	F5	Citations in method section
	F6	Citations in experiment section
	F7	Citations in discussion section
	F8	Citations in conclusion section
Cue Words Based Features	F9	Cue words for Related work citations
	F10	Cue words for Comparative citations
	F11	Cue words for Using the existing work
	F12	Cue words for Extending the existing work
Textual Features	F13	Similarity between the abstract of cited paper and text of citing paper
	F14	Author overlap -we set it to TRUE if the cited paper and citing paper share at least one common author

Further, we add six new features, including total citations received by the cited paper, similarity between the abstract of cited paper with the citation context in citing paper, and a rich set of cue words: F1 and F9 to F13.

3.2.1 Total citations received by a reference – F1. This feature counts the total number of citations received by a reference. We normalize the citation count by dividing by (2016 – y), where y is the publication year of a given reference. We use this normalization to make recent published papers comparable with the older once.

3.2.2. Number of citations to cited paper – F2. This context-based feature counts the total number of citations to the cited paper. For instance, if paper A cites paper B three times in different or the same sections then we generate one feature having value 3.

3.2.3. Number of citations per section – F3 to F8. This context-based feature counts the number of citations to the cited paper in each of six predefined sections of the citing paper. In case a section in the citing paper does not fall into one of the predefined six section types, we manually assign it to the most relevant one. Citations in the methods or experiment section are generally more indicative of work that has a significant impact upon the current work than citations in the introduction or related work sections.

For example, if a paper B has 3 citations in citing paper A - appearing 2 in related work and 1 in method section. The features will take the following values (IntroCount=0, RelatedWorkCount=2, MethodCount=1, ExperimentCount=0, DiscussionCount=0, ConclusionCount=0) for the cited paper B.

3.2.4. Cue Word Based Features – F9 to F12. There are specific cue words that scholars commonly use when referring to previous work such as "we used", "According to", "As comparing to". These cue words can be used to classify the cited work. We create a rich list of cue words for the references to classify as Related Work, Comparative, Using and Extending. These cue words are compiled from more than 80 articles discussed by Bornmann and Daniel [10] in a review paper studying citation behavior. We compile a total of 135 cue words as presented in Table 3. This feature is represented by computing cosine similarity between the cue words and the text around the reference for each class. In order to enrich our dataset, we add the synonyms for both the citation text and the cue words using WordNet. The details can be found at: https://wordnet.princeton.edu/wordnet/download/. Initially, we computed four features – one for each class – but later we opted to combine the Using and Extending class into Important and Related Work and Comparative into Incidental.

3.2.5. Similarity between abstract and reference – F13. This feature computes the similarity between the abstract of cited paper and the text around the citation in the citing paper. We use the cosine similarity measure using TF (term frequency) to compute the correspondence between cited and the citing paper. To have better measures, we apply lemmatization to obtain the root words and compute all synonyms of each word for both abstract and citation text.

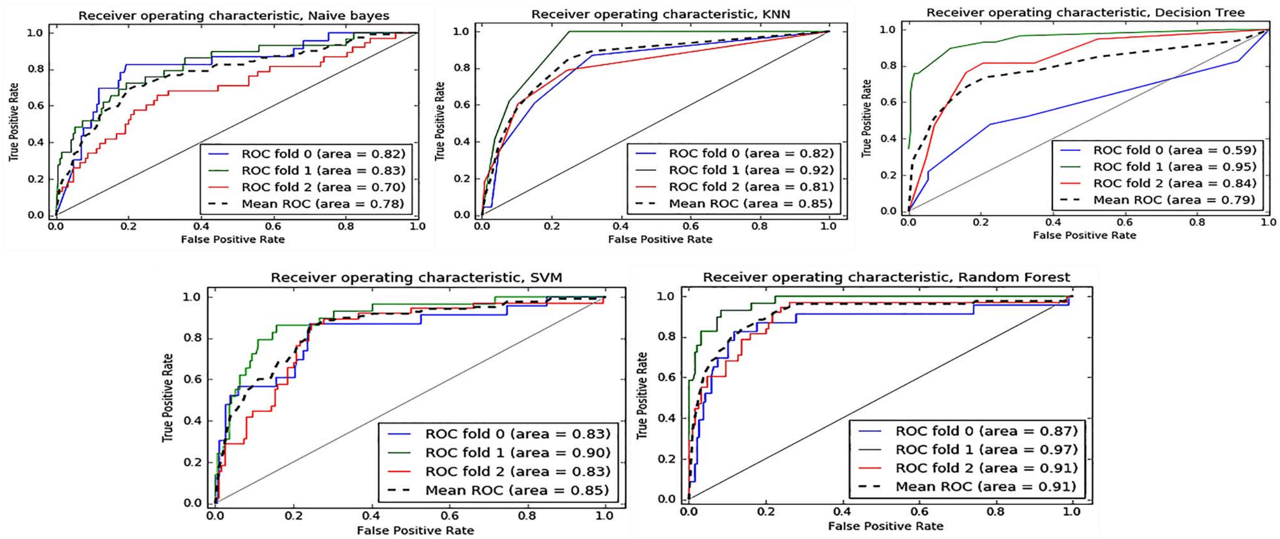

Figure 2: ROC curves of all chosen classifiers using 3-fold cross validation

3.2.6. Author overlap – F14. This feature is set it to TRUE if the cited paper and citing paper share at least one common author. This feature can help to indicate that the common author's present work may be an extension of the previous work. To compute this feature, we use only the author name. In the future, author disambiguation techniques could be used to improve the accuracy of this feature.

We normalize all features by centering on the mean and scaling to unit variance. For learning, we used five classifiers: Random forest, SVM, Naïve Bayes, Decision Tree and, K-Nearest Neighbors implemented in python using the Scikit Learn toolkit that were run using all 14 features with equal weights.

3.3 Automatic Feature Extraction

We take the following steps to automatically extract features from the dataset. To extract the text from PDF files we use XPDF in R language. To preprocess the text at different steps like tokenization; removal of punctuations, numbers, HTML tags, and stop words; POS tagging and lemmatization we use R and Python scripts. To fetch citation text, we used ParsCit [12] to segment the paper and identify citation location in full text. This includes fetching 40 words around each citation location – 20 in both forward and backward directions. Finally, we extract the title and subtitle of each section of full text paper in order to map them to one of the six predefined categories: Introduction, Related Work, Method, Experiment, Discussion, and Conclusion.

4 EXPERIMENTS

We use Receiver Operating Characteristic (ROC) curve analysis to evaluate how well our model discriminates important cited work from incidental work. Further, we use Precision-Recall curve analysis to evaluate the effectiveness of our model at different recall levels. For training the classifiers, we use three-folds cross validation technique. For Random Forest, we used Gini Index and for SVM we employ RBF kernel. Finally, using the Grid-search algorithm, we set the optimal coefficient values to train our classifiers.

4.1 Discussion of ROC curves

We employ ROC to analyze the discriminatory power of our model to distinguish between important (positive class) and incidental (negative) citations.

The important class is represented as the positive class and the incidental as the negative. Fig. 2 shows ROC curves for all five classifiers using three-fold cross validation. In addition, we also show mean ROC curves along with AUROC for each classifier. The Naïve Bayes and Decision Tree classifiers perform most poorly with mean AUC = 0.78 and 0.79, respectively. KNN and SVM classifiers show good results with mean AUC = 0.85. Random Forest performs best with AUC=0.91. These results are very encouraging for our relatively small sample size. For cross comparison, we combine mean ROC curves of all five classifiers in Fig. 3. Interestingly, the Decision Tree classifier appears to be working well for recall (True Positive Rate) less than 0.6. It actually out performs all other classifiers except Random Forest. However, for recall over 0.6 the Decision Tree shows a high false positive rate. Overall, the Random Forest classifier outperforms the other classifiers.

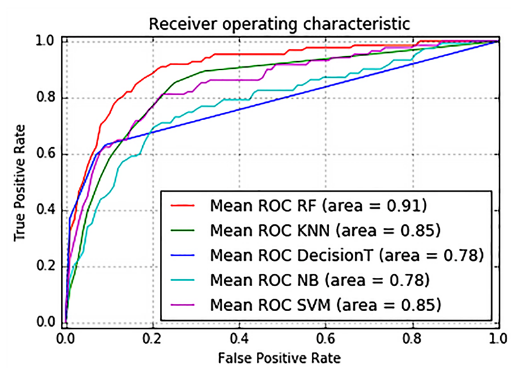

Figure 3: Mean ROC curves of all chosen classifiers

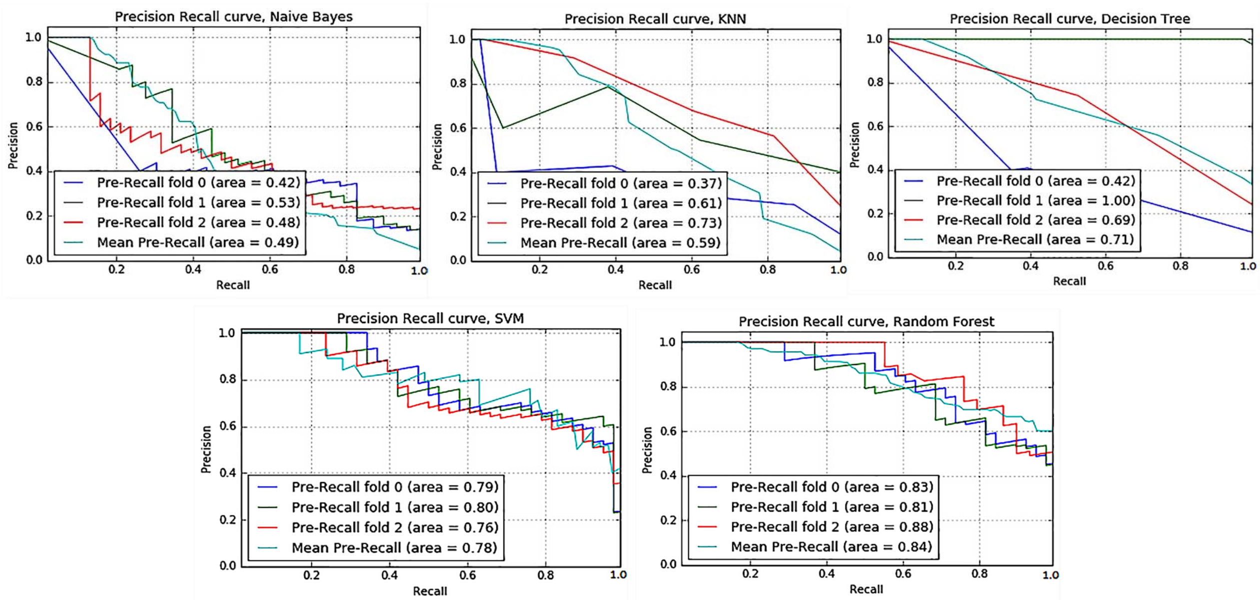

Figure 4: Precision-Recall curves of all chosen classifiers using 3-fold cross validation

While the baseline precision at 0.9 recall is under 0.2, our model with Random Forest gives above 0.65 Precision. Our model with Decision Tree outperforms KNN for recall above 0.42 while Decision Tree almost ties with SVM at 100% recall. Interestingly, our model with Random Forest gives above 0.6 precision at 100% recall, whereas the model of Valenzuela et al [1] gives around 0.35 and 0.18 precision with SVM and Random Forest classifiers respectively at the same level of recall.

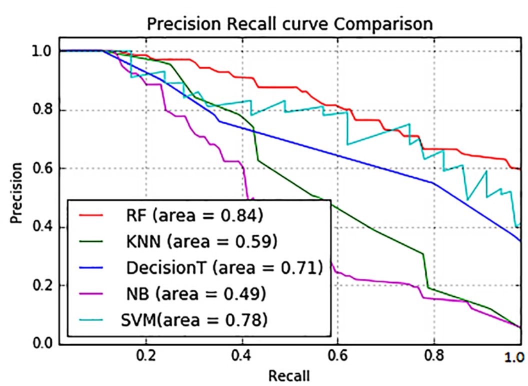

Figure 5: Mean Precision Recall curves of all chosen classifiers

4.2 Discussion on Precision-Recall Curve

While ROC analysis is one of the most commonly used metrics to evaluate binary classifiers, Precision-Recall curves can give a better sense of their performance when the data set is imbalanced [13], as in our case. Fig. 4 shows the results the five classifiers using 3-fold cross validation. We also show mean Precision Recall curves along with AUCPR for each classifier. We find that Naïve Bayes and KNN perform poorly in classifying important and incidental classes with mean AUCPR = 0.49 and 0.59, respectively. Decision Tree and SVM show good results with mean AUCPR = 0.71 and 0.78, respectively. Random Forest has an accuracy of 0.84 AUCPR, outperforming the model of Valenzuela et al [1].

For cross comparison, we combine mean AUCPR curves of all five classifiers in Fig. 5. Further, we compare these classifiers against a standard baseline that uses prior probability distribution of important label, 14.6% in our case, to randomly assign important class. We find that our model (with Random Forest, SVM and Decision Tree) performs extremely well compared to baseline model's precision.

4.3 Discussion on Learning Curve

Our classifiers exhibit different performance levels. With half of the data used for training, the Naïve Base classifier has below 70% accuracy. For the same size of training set, the Decision Tree, KNN and SVM classifiers have almost 85% accuracy. Our benchmarking model [1] achieves optimal performance (i.e. above 90% accuracy) – when half of the data is used as training examples. In contrast, our proposed model with Random Forest classifier shows optimal performance, as shown in Fig. 6, with less than 100 samples selected as training examples.

4.4 Importance of Features

We also perform a set of experiments to test the effectiveness of each feature in classification of citations. We choose the Random Forest classifier for these experimental settings since it performs best among the classifiers tested. We employ ExtraTreeClassifier to compare the importance of each feature. ExtraTreeClassifier is used to compute the feature importance also known as Gini importance or mean decrease in impurity. It is a variant of random forest. Unlike random forest, it uses the entire sample and draws random splits for each of the randomly selected features, after which the best split among those is chosen. For each tree, the importance is computed from the impurity of the splits, with higher value corresponding to more important the features.

The Table 3 shows the performance of the individual features. Our results indicate that the following seven features are most informative: Similarity between the abstract of cited paper and text of citing paper - F13, Total number of citations received by a reference - F1, Cue words for Using and Extending research - F11 & F12 (note we combine these two features as we find significant overlap of cue words used for these features), Citations in method section - F5, Cue words for Comparative citations - F10, Number of citations to cited paper - F2 and Cue words for Related work citations - F9.

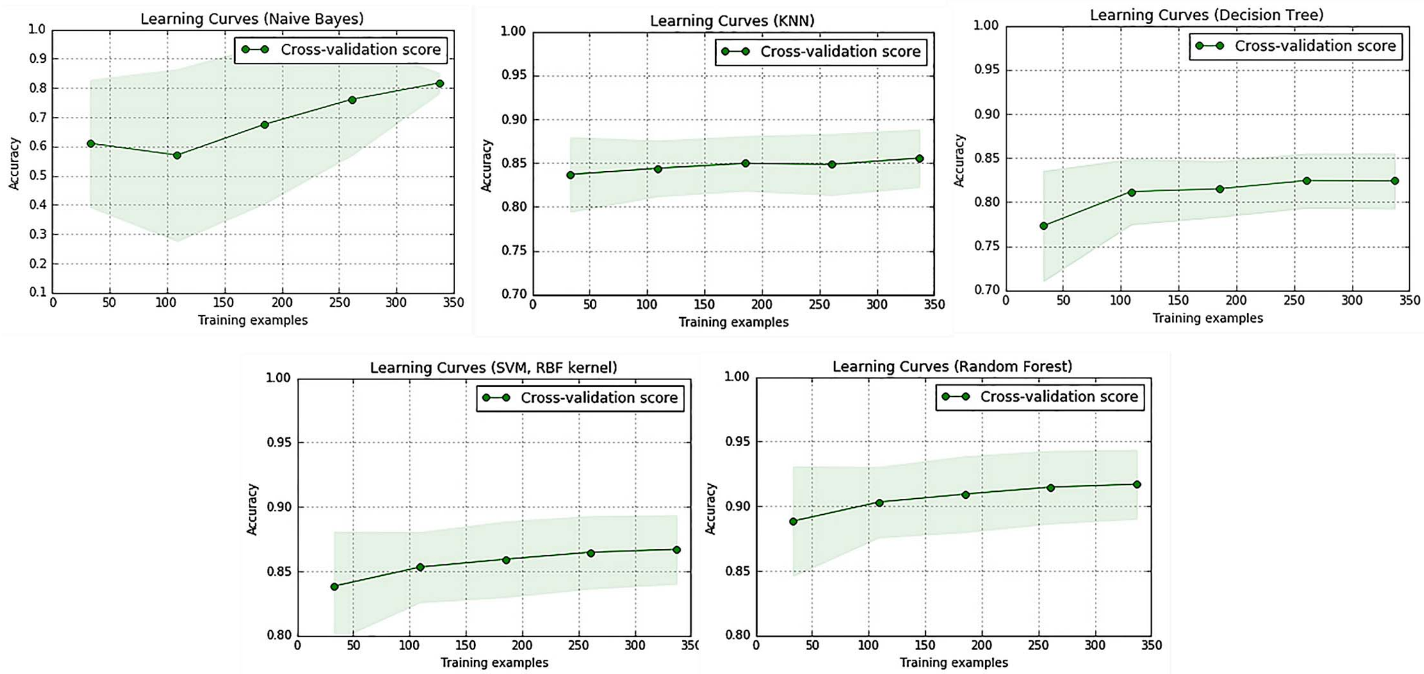

Figure 6: Learning curve of selected classifiers

Interestingly, the list of cue words appears to be very useful, since all the features have high score. In addition, we also analyze Precision-Recall curves of individual features using the Random Forest classifier.

Table 3: Extra Tree Classifier Score of Features for RF

Features	Score
Similarity between the abstract of cited paper and text of citing paper - F13	0.145
Total number of citations received by a reference - F1	0.131
Cue words for Using and Extending research - F11 & F12	0.123
Citations in method section - F5	0.123
Cue words for Comparative citations - F10	0.105
Number of citations to cited paper - F2	0.101
Cue words for Related work citations - F9	0.101
Citations in introduction section - F3	0.047
Author overlap - F14	0.040
Citations in experiment section - F6	0.035
Citations in literature review section - F4	0.021
Citations in discussion section - F7	0.017
Citations in conclusion section - F8	0.009
All Features - F1 to F14	1.0

Table 4 shows the performance of our model in terms of AUCPR when using individual features. The best performing feature according to this measure is Similarity between the abstract of cited paper and text of citing paper (F13) AUCPR: 0.64. This is followed by four features very close in performance: Cue words for Using and Extending research work (F11 & F12), Number of citation to the cited paper (F2), Citations in the introductions section (F3), and Cue words for comparative citations (F10). Note that cue word features are again prominent in this list. The least performing feature are "Citations in discussion section – F7" and "Citations in conclusion section – F8".

Table 4: AUCPR of individual feature groups using RF

Features	AUCPR
Similarity between the abstract of cited paper and text of citing paper - F13	0.64
Cue words for Using and Extending research - F11 & F12	0.55
Number of citations to cited paper – F2	0.53
Citations in introduction section – F3	0.53
Cue words for Comparative citations –F10	0.53
Cue words for Related work citations – F9	0.5
Citations in method section – F5	0.45
Total number of citations received by a reference – F1	0.43
Citations in experiment section – F6	0.41
Author overlap – F14	0.41
Citations in literature review section –F4	0.4
Citations in discussion section – F7	0.35
Citations in conclusion section – F8	0.21
All features - F1 to F14	0.84

4.5 Discussion on correlation between the features and class labels (Important vs. Non-important)

The previous analysis indicated the importance of features but not whether they were associated with important or incidental citations. To understand this, we analyze the correlation between each feature and class labels. We employ Spearman's (ρ) coefficient of correlation between the class labels and features, after normalizing them by centering on the mean and scaling to unit variance. We find a relatively high positive correlation between F2 and important class (ρ= 0.438), which confirms our intuition that an increasing number of direct citations from the citing paper to the cited paper is indicative of the importance of cited work for citing paper. Also, the number of citations in methods section is more strongly associated with important work than citations in other sections (ρ= 0.353).

We find a weak negative correlation between Total number of citations received by a reference – F1 and the important class (ρ= -0.075). So although F1 is an important feature (Table 3), higher values are associated with work that is not important. As would be expected, Citations in literature review section – F4 shows a weak negative correlation with the important class. We also measure the Pearson (r) coefficient of correlation between the features and class labels (see Table 5) which shows no significant difference from Spearman's correlation.

5 CONCLUDING REMARKS

We have introduced six new features for distinguishing important from non-important cited work. Included among the features is an extensive set of cue words (see Table 6) extracted from research papers. Experimentation with a broad set of classifiers showed random forest to perform best with our feature set. Use of these new features combined with some previously identified features [1] has enabled us to achieve an area under the precision-recall curve value of 0.84 and area under receiver operating characteristic curve 0.91. Analysis of our new features shows that all of them are among the top eight features in classification performance. Correlation analysis confirms our initial intuitions as to which features are indicative of important cited work and which are indicative of non-important cited work. A potential limitation of our work lies in the definitions of "important" and "incidental" citations. This study adopted the definitions that came with the standard dataset, which may not necessarily be accurate. Other definitions could be explored in future studies. The data and code used in this paper can be accessed at: https://github.com/slab-itu/important_citaitons_jcdl_2017.

REFERENCES

[1] M. Valenzuela, V. Ha, and O. Etzioni. 2015. Identifying meaningful citations, *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[2] X. Zhu, P. Turney, and D. Lemire. 2015. Measuring academic influence: Not all citations are equal, *Journal of the Association for Information Science and Technology* 66, 2 (2015), 408-427.

[3] B. Finney. 1979. The reference characteristics of scientific texts, *The City University of London*, London, United Kingdom.

[4] M. Garzone, and R. E. Mercer. 2000. Towards an automated citation classifier, *In Conference of the Canadian Society for Computational Studies of Intelligence,* Berlin: Springer, 337-346

[5] H. Oppenhe, N. Kando, and M. Okumura. 2000. Classification of research papers using citation links and citation types: Towards automatic review article generation, *Presented at 11th ASIS SIG/CR Classification Research Workshop*, 117-134.

[6] S. Teufel, A. Siddharthan, and D. Tidhar. 2006. Automatic classification of citation function, *In Proceedings of the 2006 conference on empirical methods in natural language processing, Association for Computational Linguistics*, 103-110.

[7] C. Dong, and U. Schafer, "Ensemble-style Self-training on Citation Classification," In *IJCNLP*, 2011, 623-631.

[8] C. Jochim, and H. Schütz. 2012. Towards a generic and flexible citation classifier based on a faceted classification scheme, *In Proceedings of the 2012 International Conference on Computational Linguistics*,1343–1358.

[9] M. J. Moravcsik, and P. Murugesan. 1975. Some results on the function and quality of citations, *Social studies of science* 5, 1 (1975), 86-92.

[10] N. Tandon, and A. Jain. 2012. Citation Context Sentiment Analysis for Structured Summarization of Research Papers, *In 35th German Conference on Artificial Intelligence*, Germany: Saarbrucken, 2012, 24-27.

[11] C. A. Sula, and M. Miller. 2014. Citations, contexts, and humanistic discourse: Toward automatic extraction and classification*, Literary and Linguistic Computing* 29, 3 (2014), 452-464.

[12] I. G. Councill, C. L. Giles, and M.-Y. Kan. 2008. ParsCit: An open-source CRF reference string parsing package, *LREC* 8, (2008), 661–667.

[13] J. Davis and M. Goadrich. 2006.The relationship between Precision-Recall and ROC curves," in *In Proceedings of the 23rd international conference on Machine learning*, ACM, 233–240.

Table 5: Spearman's correlation between the features and class labels

Features	Spearman's correlation (ρ)	Pearson correlation (r)
Number of citations to cited paper – F2	0.438	0.411
Citations in methods section – F5	0.356	0.387
Author overlap – F14	0.202	0.202
Similarity between the abstract of cited paper and text of citing paper - F13	0.201	0.197
Citations in discussion section – F7	0.145	0.163
Cue words for Comparative citations –F10	0.097	0.121
Cue words for Using and Extending research - F11 & F12	0.078	0.098
Citations in experiment section – F6	0.077	0.129
Citations in conclusion section – F8	0.057	0.057
Cue words for Related work citations – F9	0.044	0.064
Citations in introduction section – F3	0.026	0.097
Citations in literature review section –F4	-0.068	-0.013
Total number of citations received by a reference – F1	-0.075	-0.061

Table 6: Cue words for important and incidental class

Important Class			Incidental Class	
accessible in	example	resulted	although … yet	initially
according to	exhibit	resulting	although… a certain degree	less
adopt	expanded	results	another	not have
approach is/was good	explain	series	as compared	not reproduce
as described	explanation	shaded boxes	classical	notable research/work
attributed	expressed by	shown	compare	novel research/work
as attributed	extend	shown by	compared	on the other hand
bars	extending	shown in fig.	compares	originally
carried out	extension	similar	comparison	previously
characteristic	figure	similar to	currently	primary
characterized	follow	similarly	difference	rather than
considered	following	solid line	differences	recent
consists	hatched line	subjected to	different	recently
contains	hatched line	table	distinct	regardless of the validity
define	implemented	take into account	earlier	review
defined as equation	interpreted	to obtain	eliminating from	similar results for
demonstrated	investigated	use	except	than in
denoted	is known	using	except in	their agreement with
derive	is present	very close to	has been realized	unknown
derived	method … is/was good	was based on	have since been	unlikely
describe	model	was determined using	however	unusually
described-previously as	modelled by	we follow	in addition	was based on
describes	open boxes	we used	in addition to	was first
detail	proposed	written as	in common	
determined	reads	yield	in contrast	
developed	reported	yields	in the past	
examine	represent			