

VerbNet based Citation Sentiment Class Assignment using Machine Learning

Zainab Amjad¹, Imran Ihsan²
Department of Creative Technologies
Air University, Islamabad, Pakistan

Abstract—Citations are used to establish a link between articles. This intent has changed over the years, and citations are now being used as a criterion for evaluating the research work or the author and has become one of the most important criteria for granting rewards or incentives. As a result, many unethical activities related to the use of citations have emerged. That is why content-based citation sentiment analysis techniques are developed on the hypothesis that all citations are not equal. There are several pieces of research to find the sentiment of a citation, however, only a handful of techniques that have used citation sentences for this purpose. In this research, we have proposed a verb-oriented citation sentiment classification for researchers by semantically analyzing verbs within a citation text using VerbNet Ontology, natural language processing & four different machine learning algorithms. Our proposed methodology emphasizes the verb as a fundamental element of opinion. By developing and assessing the proposed methodology and according to benchmark results, the methodology can perform well while dealing with a variety of datasets. The technique has shown promising results using Support Vector Classifier.

Keywords—Citation content analysis; sentiment analysis; semantic analysis; ontology; natural language processing

I. INTRODUCTION

Sentiment Analysis is a method to categorize and recognize feelings, thoughts, ideas, or sentiments conveyed in a text, to determine the writer's intentions. Sentiment analysis depends on sentiment polarity and sentiment score [1]. Sentiment polarity [2] is the emotion expressed in a text, it can be positive, negative, or neutral, while sentiment score is based on one of the three models; Bag-of-words (BOW) model [3], part-of-speech (POS) model [4], and semantic relationships. In the Bag-of-words model, a text is described as the bag of its words, irrespective of grammar and word organization. POS tagging model identifies words in each language as one of many groups to define the role of a word. Categories of part-of-speech in the English language include nouns, adjectives, verbs, adverbs, etc. [5]. The last model is the semantic relationship, it is an association between the meanings of words.

Citation is a reference to a published source or even an unpublished one [6]. "Citation Sentiment Analysis" deals with the relationship between the citing paper and the cited paper to measure the quality of published work. Researchers usually need to analyze numerous scientific papers to find relevant articles to their work of research. Due to the significantly growing number of scientific papers, this task of analysis is

time-consuming and complicated. To resolve this issue there exists many researchers [7]–[9] who deal with the sentiment analysis of citation sentences to improve bibliometric measures. Such applications can help scholars in the period of research to identify the problems with the present approaches, unaddressed issues, and the present research gaps [10].

There are two existing approaches for Citation Sentiment Analysis: Qualitative and Quantitative [7]. Quantitative approaches consider that all citations are equally important while qualitative approaches believe that all citations are not equally important [9]. The quantitative approach uses citation count to rank a research paper [8] while the qualitative approach analyzes the nature of citation [10].

However, qualitative analysis of a citation is deeper than the simple sentiment analysis of a citation sentence. There is a need to explore the reason for a citation [9]. Charles [11] is an author of the book titled "The Informed Writer", wrote in his book "It is you who decides; what materials you need, discovers the connections between different pieces of information, evaluates the information". Thus, the author of a research paper creates a cognitive relationship between the citing paper and the cited paper while citing. Another research suggests that authors use verbs to assert their sentiment while citing another research [12], [13]. Therefore, verbs are the most important grammatical terms used in a research paper to express a stance towards another research and to provide a rhetorical context. The choice of a verb in a citing sentence plays an important role. Using Part-of-Speech tagging, it is now possible to tag verbs in a citing sentence using Natural Language Processing techniques. Combining the sentiment polarity and verbs in a citation sentence can help to understand the true nature of the author's intent.

This research aims to replace traditional citation sentiment analysis techniques by taking an ontological approach by using VerbNet Ontology and Mapping Graph [9] between verbs used within a citation to formulate opinions and its evaluation model that can identify the role of verbs in citation sentiment analysis. Section 2 describes the literature review and Section 3 has our proposed methodology. In Sections 4 and 5 experiments and results are delineated. Section 6 concludes the paper.

II. LITERATURE REVIEW

ACL Anthology Network dataset is a collection of 8736 citations from 310 research papers [10]. This sentiment corpus is a manually created dataset that can be used for automatic

classification citation sentences. In the experiments, using supervised classifiers an F-Score of 0.797 was achieved using 10-fold cross-validation. Later, a context-enhanced citation sentiment detection was performed on the same dataset [14]. In this experiment, the dominant sentiment in the citation is considered as the context that represents more than one sentiment in a citation. The effect of context windows of different lengths on the performance of a sentiment analysis system was also studied [15].

Niket Tandon and Ashish Jain [16] proposed a new technique to generate a structured summary of research papers. The proposed methodology classified citation context into one or more of five classes using a Language Model (LM) approach. Random k-Label sets with Naïve Bayes algorithm was used as the baseline to achieve 68.5% average precision. Xiaojun Wan & Fang Liu [17] used the Regression method to automatically evaluate the strength value each citation, and the strength value was used to measure the significance and influence of paper and the author. For this purpose, the Support Vector Regression method [18] was used. Bilal Hayat [19] proposed a novel automated method for the classification of citation sentiments as positive and negative. Sentiment lexicon was used to classify the citation by picking a window size of five sentences and for sentiment analysis, the Naïve Bayes classifier was used. The technique was assessed on a manually annotated dataset that consists of 150 research papers and the results depicted 80% accuracy. Cheol Kim and George R. Thoma [20] presented an automated technique to classify the sentiments articulated in Comment-on sentences using the Support Vector Machine (SVM) with a Radial Basis Kernel Function (RBF) and a Bag-of-Words input features constructed on n-grams word statistics. Jun Xu [21] presented the citation sentiment analysis of the citations in clinical research papers. For this purpose, the discussion section from 285 clinical trial papers was selected and extracted the n-grams, sentiment lexicons, and structure features. The citations were classified using Machine Learning methods and performance was evaluated using the 10-fold cross-validation method to achieve 0.8 Micro F-score and 0.719 Macro F-score.

Marco Valenzuela [22] proposed a supervised classification method that states the task of classifying meaningful citations with either two classes (important vs. non-important citation) or four classes (incidental: related work, incidental: comparison, important: using the work, important: extending the work.) Their approach used both direct citations and indirect citations. They achieved a precision of 65% for a recall of 90%. Faiza Qayyum and Muhammad Tanvir Afzal [7] presented a binary citation classification approach, using metadata-based parameters and cue-terms. Their work is close to the approach proposed by Valenzuela [22] which is the combination of metadata and content-based features, also used two types of parameters: Metadata based parameters (Titles, Authors name, Keywords, Categories, and References) and content-based parameters (Abstract and Cue-phrases). The experiments are performed on two annotated data sets, which were evaluated by using SVM, KLR, and Random Forest classifiers. The proposed model achieved 0.68 precision.

In 2018, Zehra Taskin [23] conducted a content-based citation analysis for Turkish research and they concluded that using computational linguistics for the evaluation of citation contexts provides better results. They divided the citation text into for main classes, meaning, purpose, shape, array. This research was significant for the evaluation of citation text by context. Imran Ihsan [9] proposed a Citation's Context and Reasons Ontology (CCRO) that helped to identify citations' relations using dominant verbs from citation sentences. The proposed ontology created 8 classes all extracted from Positive, Negative, and Neutral sentiments. The extracted verb was mapped to the relevant classes in CCRO based on the sentiment of the verb in a citation text. The results illustrate that the proposed ontology is reliable and complete.

VerbNet [24] is an ontology-based on Stanford Linguist Beth Levins's English Verb Classes [25]. The ontology is a lexical resource that includes both semantic and syntactic information about its contents that houses over 230 verb classes. CCRO [9] has created a knowledge-based known as "Mapping Graph" among the verbs with predicative complements in the English Language, the verbs extracted from the selected corpus using NLP and CCRO classes. Combining VerbNet Ontology and Mapping Graph proposed in CCRO, this research uses Natural Language Processing techniques to extract and map verbs within a citation sentence for semantic-based citation sentiment analysis using various machine learning algorithms.

III. PROPOSED METHODOLOGY

Fig. 1 displays the main process blocks of our proposed methodology. The methodology has four major blocks: datasets, preprocessing, feature selection, and machine learning algorithms. Details of individual blocks are.

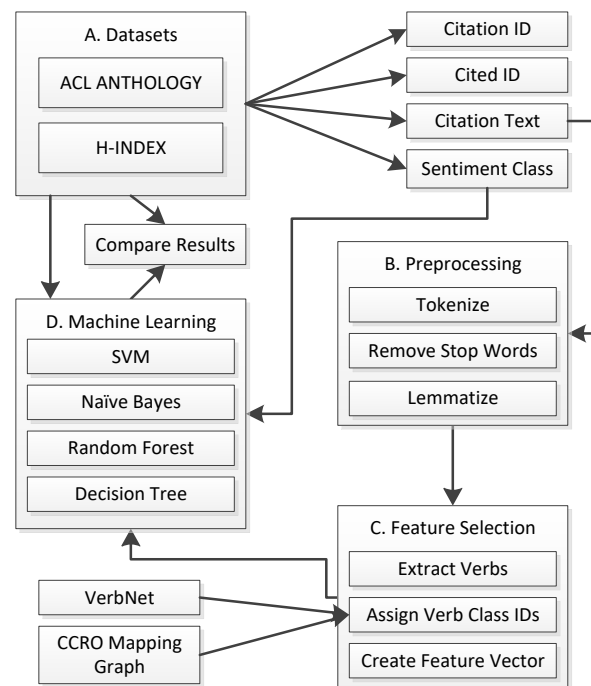


Fig. 1. Methodology.

A. Datasets

Two datasets are employed. One is the publicly available ACL Anthology Dataset while the second is the manually curated H-Index dataset. ACL Anthology Dataset comprises of all the papers published by ACL and Computational Linguistics journal. Athar [26] manually constructed a dataset comprising of 8738 citation sentences, labeled with Citing Paper ID, Cited Paper ID, Citation Sentences, and their sentiment polarity (Positive, Negative, and Neutral). The second dataset [27] is a specific version of the ANN dataset [13] comprising of 701 citation sentences with their sentiment polarity. The distribution of all three classes in both datasets is shown in Fig. 2. Kindly note, the two datasets are employed for comparative study purposes only.



Fig. 2. Sentiment Class Distribution in both Datasets.

B. Preprocessing

After selecting datasets next step is pre-processing on citation texts. The process comprises four steps. The first step is punctuation removal. Punctuation includes full stop, comma, and brackets, etc. used in writing to separate sentences and to clarify meaning. The second step is splitting up a sequence of citation text strings into pieces such as words, symbols called tokens. The third step is stop-words removal where commonly appearing words like 'is', 'a', 'an', 'it', 'which', etc. are considered as stop words and removed. The presence of stop words induces extra noise in different NLP problems that can negatively affect the results. In the last step, all the words in citing sentences are changed in their root terms. It does not simply chop off variations but uses a lexical knowledge like WordNet to gain an accurate form of words.

C. Feature Selection

In feature selection, the first step is to extract verbs from tokenized citation sentences and can be achieved using part-of-speech tagging (POS). POS is also known as grammatical tagging. This technique marks the words from a text to a specific part of speech. In our experiments, only verbs are tagged. After tagging the next step is to assign a class ID using VerbNet. The VerbNet maps the verbs to their corresponding class. It is a lexical resource that includes both semantic and syntactic information about its contents. For this mapping, a Mapping Graph is used. Using the knowledge base and the extracted verbs in each sentiment, the Mapping Graph has been formulated [9] that provides a high level of abstraction

on CCRO classes. Based on the citation context, one such property can be attributed to multiple classes. Therefore, the combination becomes a graph rather than a tree where one individual verb can belong to multiple classes based on the citation's sentiment, making the class semantically coherent.

D. Machine Learning

Based on our literature review, most of the researchers have used Support Vector Classifier (SVC), Naïve Bayes, and Random Forest Machine Learning Algorithms for the evaluation. Therefore, four algorithms were used in our proposed methodology. We have utilized the Support Vector Machine (SVM) with RBF kernel and degree 2, Naïve Bayes, Decision Tree, and Random Forest with a total no. of 10 trees and 0 maximum depth. As we have a class imbalance problem, which can lead to biasness of outcomes by always predicting the incidental class accurately. To solve this problem, we have used the SMOTE filter [28] in python. This solved the class imbalance problem by equalizing the number of classes. We have macro averaged the results of precision, recall, and F1- score.

IV. EXPERIMENTS

The experiments performed are divided into three levels. The first level of the experiment describes data preprocessing. The second level uses VerbNet Ontology to extract and map verbs from citation sentences on its class ID. The third level is to apply machine learning algorithms to classify a citation in three sentiment classes.

A. Data Preprocessing

To preprocess both datasets, a Python application was developed to remove punctuations, tokenize and remove stop-words, and lemmatization. The application using Spacy and NLTK Libraries. The resultant is a set of tokens in their base format. The sample output is shown in Fig. 3.

B. Extract Verbs

The second experiment was to extract verbs from the preprocessed citing sentences. This step was achieved using Part of Speech (POS) tagger using NLTK using algorithms from similar research [13]. The total number of unique verbs extracted from the AAL dataset was 555, and from the H-Index dataset were 337. The total occurrence of verbs in the AAL dataset was 18,789 and, in the H-Index dataset were 700. Later, these unique verbs were assigned IDs using VerbNet class IDs. Kindly note, a verb can be a part of multiple classes in VerbNet making it a graph rather than a tree. Table I shows some sample verbs and their assigned VerbNet Class ID.

C. Machine Learning Models

We have utilized the Support Vector Machine (SVM) with RBF kernel and degree 2, Naïve Bayes, Decision Tree, and Random Forest with a total no. of 10 trees and 0 maximum depth. As both datasets have a Class Imbalance problem that can lead to biases of outcomes by always predicting the incidental class accurately, SMOTE filter [28] was used. This solved the class imbalance problem by equalizing the number of classes. For the evaluation of results, macro averaged results were tabulated for precision, recall, and F1- score.

```

0 [analyze, set, article, identify, major, operation, edit, extract, sentence, include, remove, extraneous, phrase, et
ract, sentence, combine, reduce, sentence, sentence, syntactic, transformation,...]
1 [table, 3, sample, compression, compression, angle, rat, baseline, 970, 100, dt2step, 2206, 221, apade, 1009, 210,
humor, 2007, 383, table, 4, wear, rating, automatic, compression, rally, add, e...
2 [53, relate, work, discussion, trusted, model,
essentially, belong, category, work, mail, et, al. . . 1909, jing, mckinnon, 2006]
3 [1999, propose, summarization, system, base, draft, revision, jing, mckinnon, 2008, process, system, base, extraction,
cutandpaste, generation, abstractor, perform, cutandpaste, operation, jing, m...
4 [find, deletion, lead, part, occur, summary, unlike, case, jing, mckinnon, 2009]
5 [automatic, text, summarization, approach, offer, reasonably, wellperform, approximation, identify, important, sente
nce, lin, buy, 2002, schiffman, et, al, 2002, erkan, rahw, 2004, mihalcea, te...
6 [al, 1994, compression, sentence, automatic, translation, approach, knight, narco, 2008, hidden, markov, model, jing,
mckinnon, 2008, topic, signature, base, method, lin, buy, 2000, lacatuso, et, ...
7 [generally, accept, kind, postprocessing, perform, improve, final, result, shortening, fuse, revise, material, grefes
statte, 1958, mail, gate, bloomern, 1990, jing, mckinnon, 2009, harzily, et, al...

```

Fig. 3. Citation Texts after Preprocessing.

TABLE I. EXTRACTED VERBS VS VERBNET CLASS IDS

No	Verbs	Class ID's
1	Analyze	['assessment-34']
2	Set	['braid-41.2.2', 'force-59-1', 'image_impression-25.1', 'preparing-26.3-2', 'put-9.1-2']
3	Identify	['characterize-29.2-1-1']
4	Remove	['banish-10.2', 'remove-10.1']
5	Combine	['mix-22.1-1-1']
6	Reduce	['limit-76']
7	Extract	['remove-10.1']
9	Add	['mix-22.1-2']
10	Relate	['say-37.7-1']
12	Perform	['performance-26.7-1']
13	Think	['consider-29.9-2', 'wish-62']
14	Find	['declare-29.4-1-1-2', 'get-13.5.1']
15	Lead	['accompany-51.7', 'force-59']
16	Occur	['occurrence-48.3']
17	Offer	['future_having-13.3', 'reflexive_appearance-48.1.2']
18	Accept	['approve-77', 'characterize-29.2-1-1', 'obtain-13.5.2']
19	Improve	['other_cos-45.4']
20	Generate	['engender-27']
21	Neglect	['neglect-75-1-1']
22	Produce	['create-26.4', 'performance-26.7-2']
23	Observe	['conjecture-29.5-2', 'investigate-35.4', 'say-37.7-1', 'sight-30.2']

D. Performance Evaluation

1) *Classification accuracy*: Classification Accuracy is the ratio of the number of correct predictions to the total number of input examples. Classification accuracy is calculated by using the formula shown in Eq. 1.

$$Accuracy = \frac{No\ of\ Correct\ Predictions}{Total\ No\ of\ Predictions\ Made} \quad (1)$$

2) *Precision (Positive Predictive Value)*: Precision is a metric that counts the number of correct positive predictions made by the algorithm. It was calculated using the formula shown in Eq. 2.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

3) *Recall*: Recall is the metric that counts the number of correct positive predictions made from all positive predictions. It was calculated using the formula in Eq. 3.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

4) *F-Score*: F-measure combines both recall and precision into a single measure that has both the properties. Alone, neither recall nor precision expresses the complete story. So, once recall and precision have been calculated, both scores were combined into the calculation of F-measure. It is calculated by using the formula in Eq. 4.

$$F - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

V. RESULTS

After all pre-processing is applied on both datasets, the datasets are passed to the model for training. To evaluate the performance of our algorithms, the datasets were divided into two sets (Training and Validation set). 70% of the labeled data was used for training and 30 % of the labeled data was set aside for validation. After the training phase, 30% of the data was used to find out the accuracy of the algorithm. This labeled data was passed to the trained model. The model assigned labels to the verbs. These labels were then compared with the actual labels of the data. This comparison showed that our model was able to label all the verbs with an accuracy of 90%.

A. Results on AAL Dataset

Fig. 4 shows performance evaluation on AAL Dataset for four different classifiers, whereas Fig. 5 shows the precision-recall curve. The results show that SVM and Random Forest have performed better than Decision Tree and Naïve Bayes.

B. Results on H-Index Dataset

Fig. 6 shows performance evaluation on H-Index Dataset for four different classifiers, whereas Fig. 7 shows the precision-recall curve. The results show that SVM and Decision Tree have performed better than the other two.

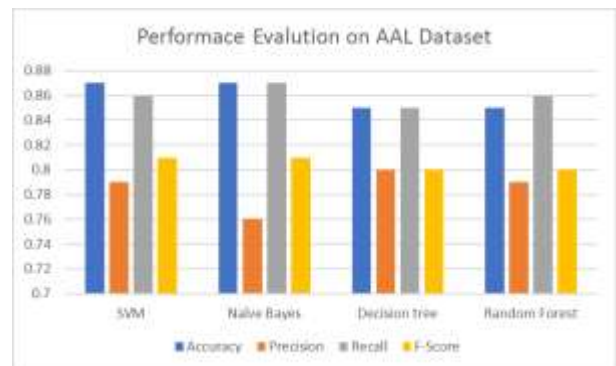


Fig. 4. Results on AAL Dataset.

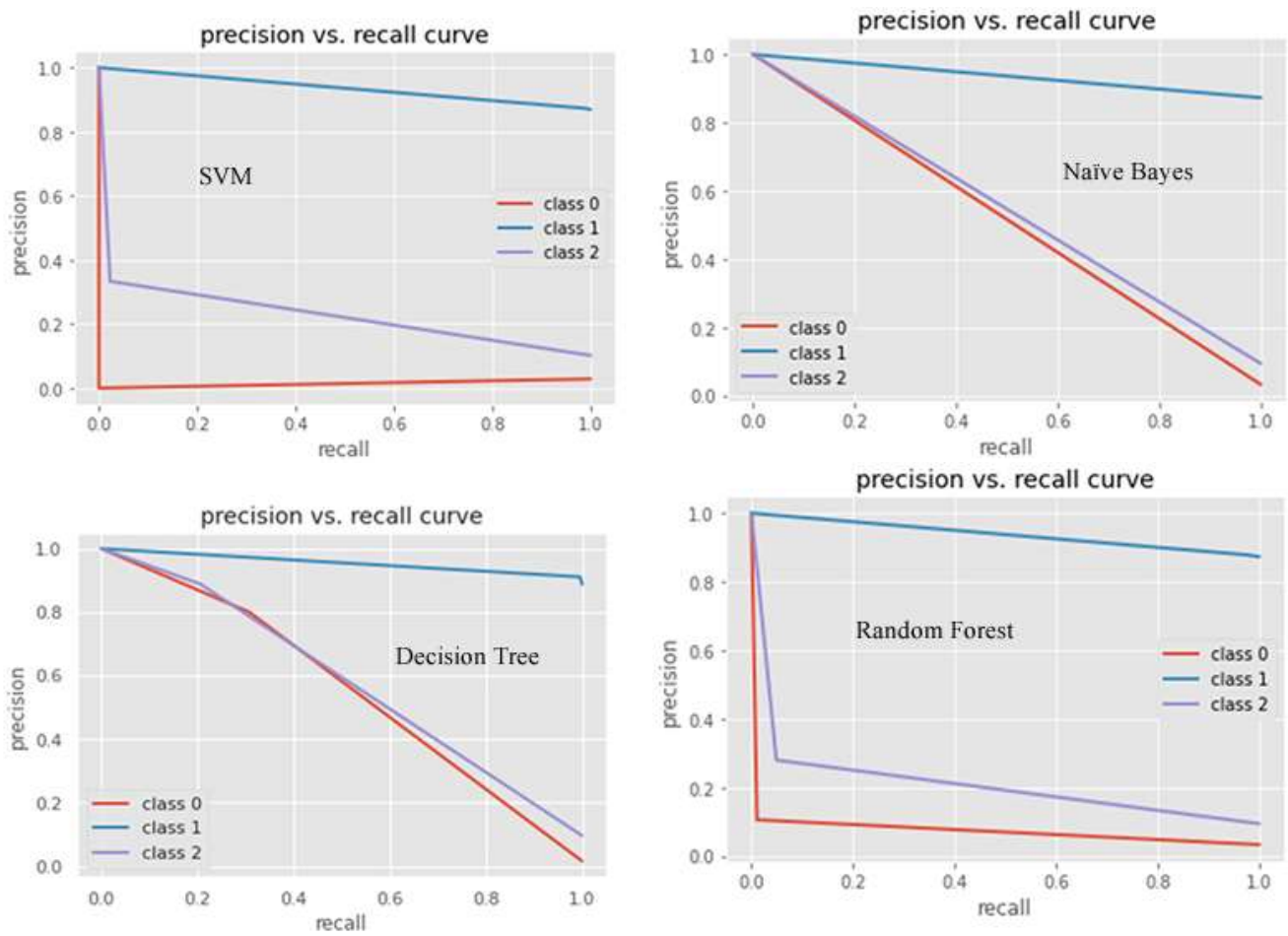


Fig. 5. Precision vs Recall Curve on 4 MLAs for AAL Dataset.

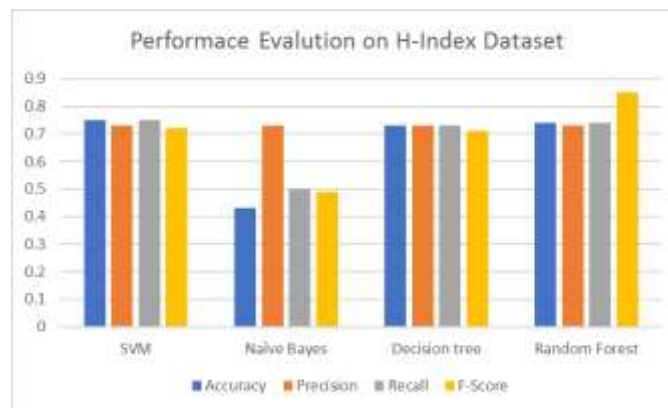


Fig. 6. Results on H-Index Dataset.

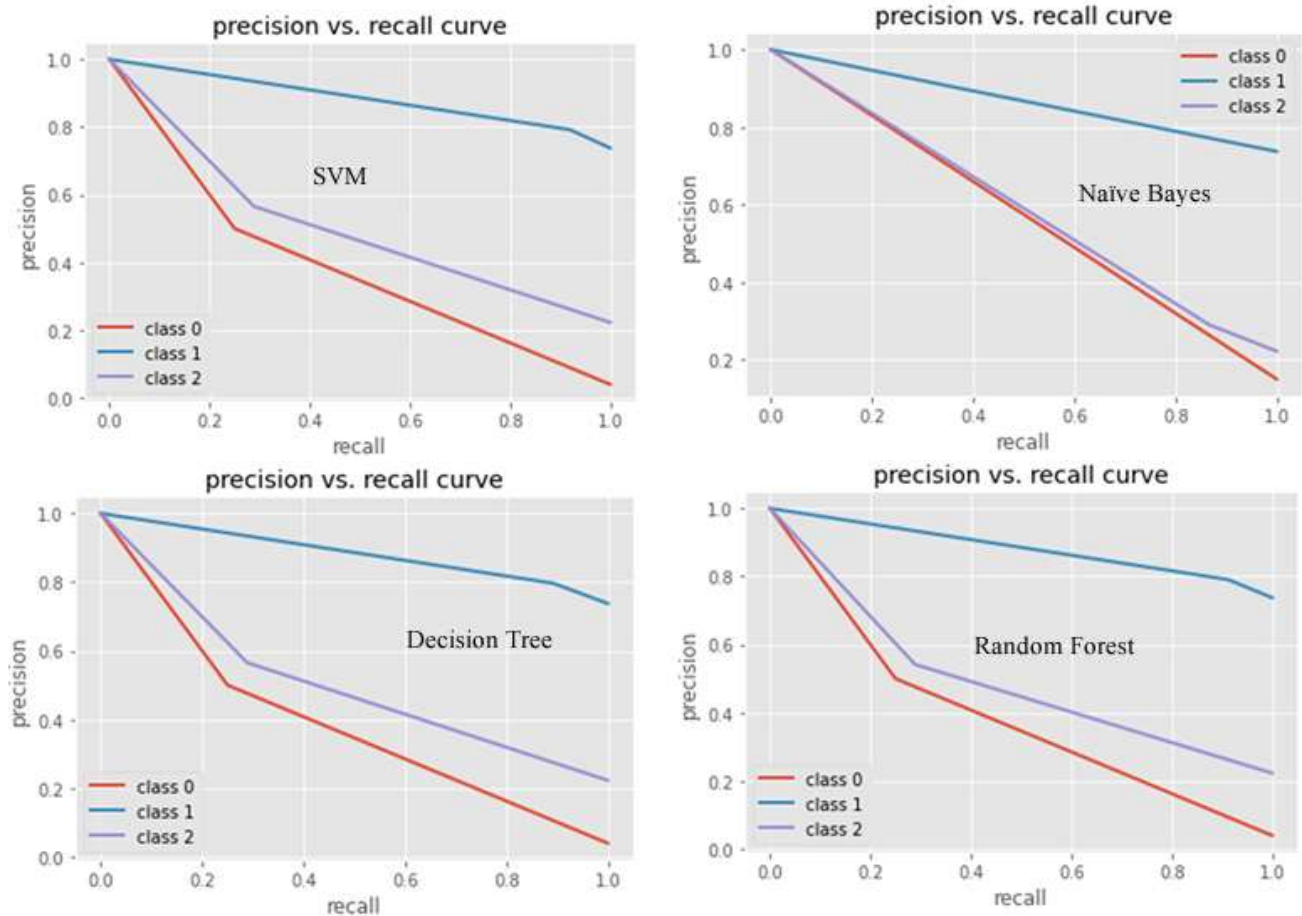


Fig. 7. Precision vs Recall Curve on 4 MLAs for H-Index Dataset.

C. Combined Results

Combined results show that SVM has given better results than Naïve Bayes, whereas Random Forest has given the best results for both datasets implying that the extracted verbs as features have shown promising results using Support Vector Classifier and Random Forest as compare to Naïve Bayes.

VI. CONCLUSION

Research is a continuous and recursive process. Every research paper and articles are built on some prior knowledge in the field. Research papers include citations to the external resources to discuss the work done by the previous researcher. With the rapid development in the research area, it becomes challenging for researchers to recognize quality research work. We have explored various existing approaches where classification methods mostly use nouns, adjectives, etc. as features. This paper proposes a new verb-based approach as an important term of opinion. We have extracted opinion structures that regard the verb as an essential component. We have used publicly available ACL Anthology Citation Dataset and our curated H-Index dataset for experiments. The experiments show 90% accuracy using Random Forests.

REFERENCES

[1] M. El-Din, "Analyzing Scientific Papers Based on Sentiment Analysis," Cairo University, 2016.

[2] B. Liu, "Sentiment analysis and opinion mining," Synth. Lect. Hum. Lang. Technol., 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[3] Y. Zhang, R. Jin, and Z. H. Zhou, "Understanding bag-of-words model: A statistical framework," Int. J. Mach. Learn. Cybern., 2010, doi: 10.1007/s13042-010-0001-0.

[4] C. N. dos Santos and R. L. Milidiú, "Part-of-speech tagging," in SpringerBriefs in Computer Science, 2012.

[5] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V. S. Subrahmanian, "Sentiment analysis: Adjectives and adverbs are better than adjectives alone," 2007.

[6] S. B. Shum, "Evolving the Web for Scientific Knowledge: First Steps Towards an ÖHCI Knowledge WebÖ TodayÖs HCI digital library," Interfaces (Providence), vol. 39, pp. 1–9, 1998.

[7] F. Qayyum and M. T. Afzal, "Identification of important citations by exploiting research articles' metadata and cue-terms from content," Scientometrics, 2019, doi: 10.1007/s11192-018-2961-x.

[8] D. W. O. Fleischmann, , An-Shou Cheng and Kenneth R., Ping Wang,Emi Ishita, "The Role of Innovation and Wealth in the Net Neutrality Debate," Commun. Inf. Lit., 2009, doi: 10.1002/asi.

[9] I. Ihsan and M. A. Qadir, "CCRO: Citation's Context Reasons Ontology," IEEE Access, vol. 7, pp. 30423–30436, 2019, doi: 10.1109/ACCESS.2019.2903450.

[10] A. Athar, "Sentiment analysis of citations using sentence structure-based features," ACL HLT 2011 - 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Student Sess., no. June, pp. 81–87, 2011, [Online]. Available: <http://dl.acm.org/citation.cfm?id=2000976.2000991>.

[11] C. Bazerman, The Informed Reader: Contemporary Issues in the Disciplines. 2010.

- [12] G. Thompson and Y. Yiyun, "Evaluation in the reporting verbs used in academic papers," *Appl. Linguist.*, vol. 12, no. 4, pp. 365–382, 1991, doi: 10.1093/applin/12.4.365.
- [13] I. Ihsan, S. Imran, O. Ahmed, and M. A. Qadir, "Sentiment Based Study of Citations Reporting Verb Corpus Using Natural Language Processing," *Corporum J. Corpus Linguist.*, vol. 2, no. 1, pp. 25–35, 2019.
- [14] A. Athar and S. Teufel, "Context-enhanced citation sentiment detection," 2012.
- [15] V. Qazvinian and D. R. Radev, "Identifying non-explicit citing sentences for citation-based summarization," 2010.
- [16] N. Tandon and A. Jain, "Citation Context Sentiment Analysis for Structured Summarization of Research Papers," in the 35th German Conference on Artificial Intelligence (KI-2012), 2012, no. i, pp. 98–102.
- [17] C. Jochim and H. Schütze, "Towards a generic and flexible citation classifier based on a faceted classification scheme," 24th Int. Conf. Comput. Linguist. - Proc. COLING 2012 Tech. Pap., no. December 2012, pp. 1343–1358, 2012, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.379.2126>.
- [18] H. F. Eid, A. Darwish, A. Ella Hassanien, and A. Abraham, "Principle components analysis and support vector machine based Intrusion Detection System," in Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications, ISDA'10, 2010, pp. 363–367, doi: 10.1109/ISDA.2010.5687239.
- [19] B. H. Butt, M. Rafi, A. Jamal, R. S. Ur Rehman, S. M. Z. Alam, and M. B. Alam, "Classification of research citations (CRC)," in CEUR Workshop Proceedings, 2015, vol. 1384, no. January, pp. 18–27.
- [20] I. C. Kim and G. R. Thoma, "Automated classification of author's sentiments in citation using machine learning techniques: A preliminary study," 2015, doi: 10.1109/CIBCB.2015.7300319.
- [21] J. Xu, Y. Zhang, Y. Wu, J. Wang, X. Dong, and H. Xu, "Citation Sentiment Analysis in Clinical Trial Papers," *AMIA ... Annu. Symp. proceedings. AMIA Symp.*, vol. 2015, pp. 1334–1341, 2015.
- [22] M. Valenzuela, V. Ha, and O. Etzioni, "Identifying meaningful citations," *AAAI Work. - Tech. Rep.*, vol. WS-15-13, pp. 21–26, 2015, [Online]. Available: <http://ai2-website.s3.amazonaws.com/publications/ValenzuelaHaMeaningfulCitations.pdf>.
- [23] Z. Taşkın and U. Al, "A content-based citation analysis study based on text categorization," *Scientometrics*, vol. 114, no. 1, pp. 335–357, 2018, doi: 10.1007/s11192-017-2560-2.
- [24] A. M. Giuglea and A. Moschitti, "Semantic role labeling via FrameNet, VerbNet and PropBank," 2006, doi: 10.3115/1220175.1220292.
- [25] B. Levin, *Book reviews*, vol. 37, no. 3. University of Chicago Press, 2012.
- [26] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, "The ACL anthology network corpus," *Lang. Resour. Eval.*, vol. 47, no. 4, pp. 919–944, 2013, doi: 10.1007/s10579-012-9211-2.
- [27] I. Ihsan and M. A. Qadir, "CCRO: H-Index Dataset," 2020. <https://github.com/imranihsan/CCRO/blob/master/CCRO-CC2.csv>.
- [28] J. A. Sáez, J. Luengo, J. Stefanowski, and F. Herrera, "SMOTE-IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering," *Inf. Sci. (Ny)*, 2015, doi: 10.1016/j.ins.2014.08.051.