

A New Scheme for Citation Classification based on Convolutional Neural Networks

Khadija Bakhti¹, Zhendong Niu^{1,2}, Ally S. Nyamawe¹

¹*School of Computer Science and Technology*

Beijing Institute of Technology

Beijing, China

²*School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA*

(bakhti.khadidja, zniu)@bit.edu.cn, nyamawe@udom.ac.tz

Abstract—Automated classification of citation function in scientific text is a new emerging research topic inspired by traditional citation analysis in applied linguistic and scientometric fields. The aim is to classify citations in scholarly publication in order to identify author’s purpose or motivation for quoting or citing a particular paper. Several citation schemes have been proposed to classify the citations into different functions. However, it is extremely challenging to find standard scheme to classify citations, and some of the proposed schemes have similar functions. Moreover, most of previous studies mainly used classical machine learning methods such as support vector machine and neural networks with a number of manually created features. These features are incomplete and suffer from time-consuming and error prone weakness. To address these problems, we present a new citation scheme with less functions and propose a deep learning model for classification. The citation sentences and author’s information were fed to convolutional neural networks to build citation and author representations. A corpus was built using the proposed scheme and a number of experiments were carried out to assess the model. Experimental results have shown that the proposed approach outperforms the existing methods in term of accuracy, precision and recall.

Index Terms—Citation Annotation, Citation Scheme, Deep Neural Networks, Citation Function Classification, Convolutional Neural Networks.

I. INTRODUCTION

In the previous published research works the citation is categorized as a tool to calculate impact factor with an objective to know how the citation is used [1], [2]. Citation function classification is defined as the reason or motivation that why the authors cite others works in their literature, and the field of research concerned with classifying citations into classes based on the purpose behind the citations. Classification of citations could provide precise representation of the influence or the impact of a publication. For example, by considering only citations that are important to the citing paper and discarding citations that are perfunctory. The first step in citation function classification includes selecting a number of functions that citations can be categorized into, which is called a citation function classification scheme. When a scheme has been selected, a classification method is used to carry out the classification of citations.

Several citation function classification schemes have been created with a different number of functions and levels of granularity [3]. For example, [4] established a citation

scheme containing four dimensions with two functions in each dimension. Each dimension groups two related classes together; a citation can belong to one class only from one or more dimensions. Different names are used in the literature to represent specific purposes for citations such as “category”, “class”, “type”, “reason” and “facet”. We refer to the different names throughout the paper by the word “function”.

Manual citation function classification has been proposed, but subsequently automated classification became inevitable due to the large number of publications produced on a daily basis [3]. Automated citation function classification has been carried in the literature into two ways; the first way is the use of rule based methods where domain experts developed rules that were coded into computer programs to perform citation function classification [5]. The rules were created based on a set of human labeled citations where each citation was labeled with a function or label revealing the related purpose. The second way involves applying supervised machine learning techniques [6] where a set of citations were labeled by human annotators to build the training phase.

Previous studies on automated citation function classification commonly used rule-based and supervised machine learning methods [5], [7]. However, the rule-based techniques do not generalize well for citations that have never been seen by the domain experts. Therefore, multiple schemes have been proposed with different granularity varies from 35 to 3 functions [8]. However, there is no standard scheme established for citation function. Therefore, there is no way that a scheme can allow authors to frame their citations and how this framing can influence the use by future citers. [7] proposed a citation scheme for classifying citation function into six functions namely based on/supply, useful, weakness, contrast, acknowledge, hedges. [9] proposed a new scheme to annotate the citations which has seven functions: background, motivation, uses, extension, continuation, comparison, and future. Regarding these proposed schemes, there is no defined standard for citation classification schemes. However, the majority of functions like based on/useful and uses/extension have the same purpose; and similarities exist between functions could be difficult for an annotator to differentiate them for the future use. Moreover the usability of the functions proposed is limited and cannot be adopted for all the annotators from

different domains. Therefore, in this context we unify the functional common roles in several classifications and grouped together all similar functions in some categories which reflect the particular reason or motivation a citation is serving in the discourse. By focusing on the dimension of organic or perfunctory citations following [4] scheme, we divide the citations into five general functions. For this initial study, we have limited to use only the top level functions. We propose these functions mainly due to the following reasons: first of all, these proposed functions cover the most general and mutually exclusive citation functions for different domains; one could facilitate the annotation because it will be easy for annotators to have these functions separated and easily to use them later on. Secondly, it is easy to depict a typical scientific publication based on citations from these functions. Thus, our proposed strategy is valuable for the construction of further detailed citation function classification models with more refined functions. The proposed functions are: *Useful*, *Correct/Weakness*, *Contrast*, *Mathematical* and *Neutral*. The proposed approach can solve the limitations of supervised learning approaches such as the incomparable citation schemes used to label the training sets provided to the supervised algorithms and the high cost of annotating the training sets by humans.

Regarding the success of neural network methods in document classification models, [10] proposed a model for designing features from words representation in neural network. The citation function classification has recently become an active field to design new features for citation function classification based on neural network methods by identifying author's reasons for citing the literature. A deep learning approach based on convolutional neural network (CNN) algorithm with a specific layer for author information is proposed in our case to learn author's information (including author-id and name) for embedding word vector in the input layer. We will demonstrate that the proposed approach using CNN is able to solve the feature selection and representation issues automatically and achieve better results compared to other existing methods in citation function classification task. The model is tested on our corpus based on the proposed scheme and the output vector is classified into the proposed five functions.

The paper is structured as follows: section II responds to the state of the art on citation function classification. In section III, the proposed methodology is presented. In section IV and V, experiments and results are discussed. Finally, a conclusion and future work are drawn in section VI.

II. STATE OF THE ART ON CITATION FUNCTION CLASSIFICATION

Citation function classification has been defined as the process of identifying the functions or purpose of quoting from other works [11]. In other words, it means the process of detecting the relationship between citing paper and cited paper. Many authors formulate this relationship as a citation schemes and reported several schemes to identify the

influence of cited papers on citing papers. For example, the citation scheme proposed in [12], where fifteen reasons were suggested to justify the quoting from previous work. Another citation function classification contains four dimensions which include conceptual or operational use, evolutionary, organic or perfunctory and confirmative or negation [4]. [13] reported seven argumentative zones as another classification scheme, namely: background, other, own, aim, textual, contrast, and basis according to the citation role in the author's argument. However, this scheme has some limitations, due to the large number of functions used which is time-consuming for citation annotation and cannot process a large data set of documents.

The automatic classification of citation functions can handle the schemes limitations. [14] used rule-based approach with cue words to reduce the citation functions into three categories: reference type B, C, and O. Another classification of the citation was proposed by [3] where they chose 116 articles in a random way from the Computation and Language E-Print Archive and classified twelve citation functions into four categories. In [8], authors used a semi-supervised learning method and the Naive Bayes (NB) as the main technique with features such as negation and cue words.

In [15], based on the implementation of the hybrid method algorithm, authors used discourse as their tree model and analyzed part of speech to find out citation relations regarding contrast and corroboration. [16] proposed an unsupervised bootstrapping algorithm, which led to the categorization of two concepts: application and technique. [8] used the Support Vector Machine (SVM) algorithm with linear kernel method and established a faceted classification of citation links in networks that are functional, perfunctory, and ambiguous cases. In [11], authors classified the purpose and polarity of a citation using the SVM algorithm, along with the trained classification model and linear kernel to the ACL Anthology.

Recently, citation function scheme has been created using clustering techniques found to be useful to address the annotation difficulty in previous schemes. [17] proposed an approach based on contains semantic and syntactic-based models. Their models employed multiple similarity methods to calculate the similarity between citations sentences, each cluster of similar sentences considered as a citation function.

Moreover, authors in [18] explored these issues by selecting the relevant verb in a citation sentence. They labeled each citation sentence by using semantic role labeling and then proposed six rules to extract and select the best verb that represents the citation sentence. Their rules were evaluated using four test sets and their results are reasonable. To this end, we propose here a different concept based on the deep neural networks model to address citation function classification task.

III. METHOD

In this section, we first introduce our methodology for dataset creation and then describe the CNN based model for citation function classification. Each step will be presented below.

A. Dataset selection

The citation corpus was built from ACL Anthology network ¹(ANN). ACL Anthology is an academic repository that contains full-text articles with associated meta-data. Hundred papers were chosen as a corpus size for citation sentences extraction. Our choice of these papers follows a number of previous works [6], [7]. We have used parsing rules to extract citation sentences followed by regular expressions for data cleaning, and non-citation sentences were excluded. After this step, 8700 sentences were obtained and passed as inputs for citation annotation process. As expected from previous works [7], [9], we found some citation functions were infrequent. Therefore, we attempted to build frequent functions *mathematical*, *correct*, *follow* and *neutral*, by basing on keywords used towards extracting citing citations. For example, the word *statistical* for the *mathematical* function. In the citation annotation process, three PhD students worked separately as annotators to manually annotate the citation sentences using the proposed scheme. The annotators not only focused on labeling word in the citation sentence, but also read the entire sentence and the whole context where the sentence is located. Then made a decision on the function of the citation and determine function of each citation by choosing from the five functions described in Table I. To test annotation reliability, we measured inter-annotator agreement between three annotators, we used κ coefficient as proposed in [19]. We used a small section of the corpus about 800 citations to analyze them according to their function. Inter-annotator agreement was $\kappa = 0.76$ with parameter $n = 5$ and $N = 800$. The results is quite high given the fact that Kappa value of 0.76 is considered as stable [3]. Table II illustrates the distribution of the dataset.

B. Model architecture

Fig.1 shows our proposed model architecture based on CNN method. In this model, we have a citation sentence consisting of n words as $\{w_1, w_2, w_3, w_4, w_i, \dots, w_n\}$ where w_i means the i^{th} word in the sentence $\in R^d$ and the d is dimensional word vector corresponding to the word. The output of word

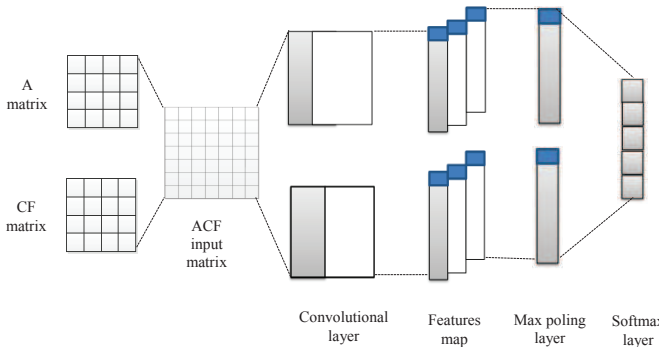


Fig. 1. Architecture of citation function classification based on CNN model.

¹<http://clair.eecs.umich.edu/aan/index.php>

TABLE I
THE PROPOSED CITATION FUNCTIONS FOR CITATION CLASSIFICATION.

Functions	Description
useful	The citation sentence is classified as use if the cited work utilized or followed data, method, tools from the citing work.
contrast	Is reserved for correct of previous research, addressing by the authors such as error, weakness from cited paper.
mathematical	Describe the comparison between the cited own work and other works, the result can be positive or negative.
correct	The cited work base on tools, results, statistical tables, algorithms from citing paper.
neutral	Expression of author using own language marked as no useful interpolation or the description of specific method, concept.

TABLE II
THE DISTRIBUTION OF THE DATASET.

Functions	The number of citation sentences	Ratio
useful	2195	24.81%
contrast	1800	20.68%
mathematical	1846	21.21%
correct	1700	19.54%
neutral	1195	13.73%
Total Sentences	8700	100%

vectors gets a real-valued vector known as word embedding. The word embedding is a powerful technique to capture the semantic and syntactic of words and also it could be useful to extract features from the sentence. Therefore, following [10], [20] strategy of word vectors representation, we used word2vec to build our CF matrix as depicted in Fig.1. In addition, we proposed to exploit author's information such as personal demographic data (id-author, name) to build a matrix A. The model takes as inputs the citation sentences and the author that cites the paper with relative information. Then we link all the authors with their citation sentences, which have the same function from our functions.

In our model, Author Citation Function (ACF) is a matrix that combines the representation of citation sentences and the author's information as presented in equation (1).

$$ACF = A \cdot CF \quad (1)$$

Finally, the ACF is passed to the CNN method to classify citation into our proposed functions.

In the CNN, the convolutional process consists of applying filters $W \in R^{h \times d}$ in a window of h words in the sentence of length n $\{w_{1:h}, w_{2:h+1}, \dots, w_{n-h+1:n}\}$. We have chosen multiple convolutional filters with varying filters window size from 3 to 5, and applying these filters using non-linear activation function (In our model we used wide Rectified Linear Unit (Relu) as the activation function, [21]) for each window of words within the citation sentence to produce a new feature p_i of size $n - h + 1$. A feature p_i is generated from a window of words $w_{i:i+h-1}$. Let consider the following ex-

ample illustrating the non-linear activation function operation given as:

$$p_i = f(W \cdot w_{i:i+h-1} + b) \quad (2)$$

where $b \in R$ is the bias, f is the non-linear activation function. The max-pooling operation, [22] is then applied. We used max-pooling because it is widely used, and the idea is to take the maximum value p_{max} from the feature map as the most important feature among one map P .

$$p_{max} = \max\{P\} = \max\{p_1, \dots, p_{n-h+1}\} \quad (3)$$

C. Function classification

To perform citation function classification, our classifier used citations with functions. The performance of the classifier can be affected by over-fitting problem, which could come from the weakness of the neural net. We have employed the dropout regularization to prevent over-fitting problem of the hidden units in the classifier. In the classification stage, we feed the final feature map to the softmax layer. We chose the softmax because it is commonly used for classification problem, which gives a probability of the sample belongs to each label (class). The outputs of softmax layer can be interpreted as conditional probabilities. Equation (4) shows the softmax function formula.

$$Softmax_i = \frac{e^{x_i}}{\sum_{j=1}^L e^{x_j}} \quad (4)$$

where L is number of labels (we have five functions as labelling (classes)) and x_j is the weight vector of the L^{th} label.

IV. EXPERIMENTS

Experiments were carried out using the dataset described in section III.A in order to test our approach by applying it into classification of citation function. The results from the experiments were then compared.

A. Experimental Setting

First, let address some hyper-parameters used. For each filter size as a window, we chose 3, 4, 5 respectively. We enable the dropout in training and disable it in evaluating the model and set 0.5 as a dropout rate. We used 10-fold cross validation for training and evaluating our model. We also apply the loss function classifier to correct and minimize errors that our network makes [23]. For evaluating the model, we calculate the accuracy using the standard accuracy formula described in [24].

B. Baseline methods

We compared the proposed ACFNN model with the following state of the art methods for citation function classification:

- N-gram+SVM: this method uses n-gram and train classifier with SVM [7].
- Word2vec+SVM: this method considers each function as a separate feature and train classifier with SVM [9].
- Word2vec +Naive Bayes: this method uses vector and train classifier with Naive Bayes [9].

TABLE III
ILLUSTRATION OF ACCURACY MEASURE.

The method	Acc(%)
N-gram+SVM (Hernandez)	56.8
Word2vec+SVM (Jurgens)	57.1
Word2vec+Naive bayes (Jurgens)	55.4
Cue phrases+IBK algorithm (Toufel)	52.2
CNN (no A)	58.2
ACFNN	62.7

- Cue phrases or meta-discourse: this method uses cue phrases and train classifier with IBK algorithm [3].
- CNN (no A): in this method, we remove matrix based representations (A) from ACF and train the model with CNN.

C. Results and discussion

In this sub-section, we report the details of the experimental results as presented in the following.

1) *Accuracy measure*: The results of the experiments are presented in Table III. As output, the final vector is a fine grained classification into five functions: *useful*, *contrast*, *mathematical*, *correct* and *neutral*.

Comparing the results of the proposed model tested along with the baseline methods, the results indicate the performance calculated in accuracy (Acc) by incorporating different feature sets with the batch size of 25, obtaining 62.7% as a best performance achieved using the proposed (ACFNN) model. Comparing (ACFNN) with CNN (no A), we can clearly observe that our model achieves the highest accuracy (62.7%), followed by the CNN (no A), which has an accuracy of 58.2%. The results show that the proposed model improves the classification accuracy by 4.5%, and this illustrates that author’s information can improve the impact of the importance of their integration in the citation function process to handle the problem of citation function classification task.

The baseline’s approaches n-gram features (SVM), word vector features (SVM, Naive Bayes), and cue phrases (IBK algorithm) suggest that word2vec features with deep neural networks models (word2vec+CNN) dual an improvement to oriented methods for a better classifier for citation function. Furthermore, the capacity in exploiting information is proven by the CNN workflow by scanning the combination of words sequentially and retaining the sequential information to attract a pool operation, which can bridge the information space at both ends of the citation sentence. Thus, it is evident that the CNN can handle the problem of manual features extraction.

2) *Performance evaluation*: Experiments were conducted to evaluate the performance of the proposed ACFNN. In comparing the performance of the proposed ACFNN approach against the baseline methods, we used precision, recall and f-measure metrics. In using precision and recall evaluation metrics, labels are mapped into a binary scale (relevant versus not relevant). We also considered learning elements as “not relevant/not classified” and “relevant/classified”. The description of precision and recall metrics is shown in Table IV. Precision

TABLE IV
PRECISION AND RECALL METRICS.

	Classified	Not Classified
Used	True Positive (tp)	False Negative (fn)
Not Used	False Positive (fp)	True Negative (tn)

is the ratio of relevant instances selected by the classifier to the number of instances selected. A learning element is considered non-relevant if the classifier ignores it.

$$Precision = \frac{\text{Correctly classified instances}}{\text{Total classified}} = \frac{tp}{tp + fp} \quad (5)$$

Recall is the ratio of relevant instances selected to the number of relevant instances.

$$Recall = \frac{\text{Correctly classified instances}}{\text{Relevant instances}} = \frac{tp}{tp + fn} \quad (6)$$

where relevant instances is the number of learning elements classified as relevant by the classifier.

F-measure is the harmonic mean of precision and recall, F-measure uses both precision and recall to correctly assess the efficacy of the classification.

$$F - \text{measure} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

The results of citation functions classification in 10-fold cross-validation are given in Table V. The results are conducted in three overall measures: Precision, Recall and F-measure of five functions. Precision for all the functions is above 0.58. To test the contribution and success of the proposed functions, we used Macro-F which is the mean average of F-measure of all five functions. In the case of Macro-F, regarding the reported result; we can see that the classification yield higher values in the functions such as *useful* and *mathematical* than other functions. The distribution of the citations functions as shown in Table II are: 24.81% *useful*, 20.68% *contrast*, 21.21% *mathematical*, 19.54% *correct* and 13.73% *neutral*. We found that total number of *useful* and *mathematical* citations is higher than the other citations, this empirically confirm that the authors are likely to use, follow or extend (*useful*) works from the cited works ($p=0.59$) as well as they are more focusing on mathematical concept (*mathematical*) such as methods, statistical tables, results from the citing work ($p=0.59$). In the *contrast* function, as we know that the authors start the state of the art with an objective (compare) the previous works. In addition to this, in the *correct*, the authors address the errors and weakness of previous works and suggest solutions to correct them, as shown in the Table V ($p=0.58$). Finally, the

TABLE V
RESULTS OF CITATION FUNCTION CLASSIFICATION.

	Precision	Recall	F-measure
useful	0.62	0.59	0.60
contrast	0.58	0.57	0.57
mathematical	0.61	0.58	0.59
correct	0.60	0.57	0.58
neutral	0.58	0.56	0.57
Macro-F = 0.58			

citations which do not belong to any of the above citations are tag as not useful description (*neutral*). The analysis of the functions indicates that there is higher negative correlation between all functions, this leads to a conclusion that these proposed functions outperform the state-of-art citation schemes described in section I (no similarities between them). Thus, it is evident that the proposed five functions can cover the most general functions and increase the performance of the classification task. Fig.2 illustrates the performance of the proposed model in terms of precision into batch size of 25 in comparison with the baseline methods. The experiment was repeated in different number of iterations.

Comparing results from Fig.2, it is observed that the proposed model provides best performance in terms of precision than the other methods with batch size of 25 regardless of the number of iterations.

It is evident that as shown in Fig.3 the proposed model outperforms the baseline methods in terms of recall. The experiment was repeated for the different number of iterations.

V. DISCUSSION

The results conducted using our corpus have shown the effectiveness of the proposed model. We can see that our model (ACFNN) significantly outperforms existing methods such as SVM, Naive Bayes and traditional CNN in terms of accuracy, precision and recall. The reason is that since the CNN can absolutely capture the semantic content of citations and select number of features needed automatically. In addition, using the concept of word embedding, we can acquire richer features automatically. Our results suggest that word2vec+CNN concentrate on weakness of prior works and had the largest impact on the performance. With the help of the authors information in concatenation with the citations and feed them to CNN, making our model more efficient and outperforms the baseline CNN by 4.5% in accuracy. As we stated the limitations of previous works in section II, there is no standard scheme up to date and it is difficult to distinguish between the functions with close similarities. Regarding the high frequency of usage frequent functions shown in Table II, we believe that our proposed scheme can handle the problem

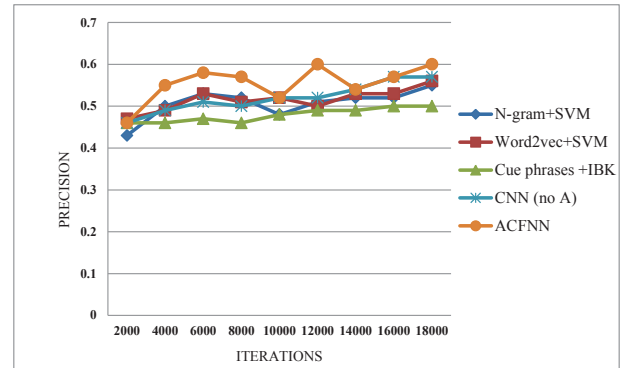


Fig. 2. The performance of the methods in term of precision.

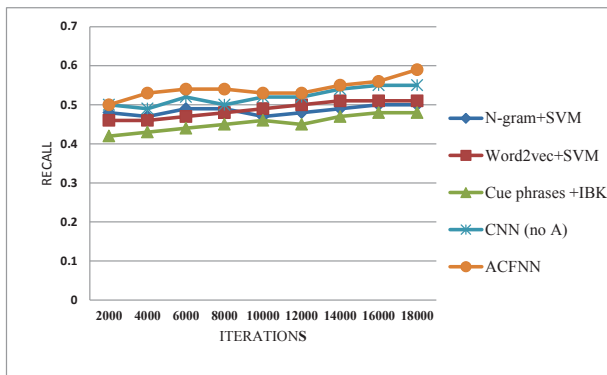


Fig. 3. The performance of the methods in term of recall.

of annotation for citation function. In addition, these functions can cover the most general and mutually exclusive citation functions for different domains. Moreover, these functions remain an important line for the future use since will be easy for the annotators to separate them later.

VI. CONCLUSION

In this paper, we have presented an approach that uses the CNN based model combined with author's information to classify citation sentences into five functions. Experimental results show that the proposed method is able to identify authors' reasons semantically. Moreover, combining citations with authors' information achieves best performance in our corpus. Therefore, our proposed scheme is able to handle the weakness of the citations annotation and can be used in different domains. The proposed model reveals that CNN can outperform the shallow classification for citation function classification task. Valuable information can be extracted using the data citation function, which will have a real interest to help in the search of high-quality papers. Our future work is to explore other deep learning approaches such as Long Short-Term Memory Networks (LSTM) which is a type of recurrent neural network (RNN).

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61370137), the Ministry of Education-China Mobile Research Foundation Project (No. 2016/2-7) and the 111 Project of Beijing Institute of Technology.

REFERENCES

- [1] Jingqiang Chen and Hai Zhuge. Summarization of scientific documents by detecting common facts in citations. *Future Generation Comp. Syst.*, 32:246–252, 2014. 10.1016/j.future.2013.07.018.
- [2] Abdallah Yousif, Zhendong Niu, John K Tarus, and Arshad Ahmad. A survey on sentiment analysis of scientific citations. *Artificial Intelligence Review*, pages 1–34, 2017.
- [3] Simone Teufel, Advait Siddharthan, and Dan Tidhar. Automatic classification of citation function. In *EMNLP 2007, Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 22-23 July 2006, Australia*, pages 103–110

- [4] Michael J Moravcsik and Poovanalngam Murugesan. Some results on the function and quality of citations. volume 5, pages 86–92. CA,1975.
- [5] Mohammad Abdullatif. Making the h-index more relevant: A step towards standard classes for citation classification. In *Workshops Proceedings of the 29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013*, pages 330–333, 2013. 10.1109/ICDEW.2013.6547476.
- [6] Myriam Hernández Alvarez and José M. Gómez. Citation impact categorization: For scientific literature. In *18th IEEE International Conference on Computational Science and Engineering, CSE 2015, Portugal, October 21-23, 2015*, pages 307–313.
- [7] Myriam Hernández Álvarez, José M Gómez, Patricio Martínez-Barco, et al. Annotated corpus for citation context analysis. 2016.
- [8] Han Xu, Eric Martin, and Ashesh Mahidadia. Using heterogeneous features for scientific citation classification. In *Proceedings of the 13th conference of the Pacific Association for Computational Linguistics*, 2013.
- [9] David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. Citation classification for behavioral analysis of a scientific field. 2016.
- [10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [11] Amjad Abu-Jbara, Jefferson Ezra, and Dragomir R. Radev. Purpose and polarity of citation: Towards nlp-based bibliometrics. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, USA*, pages 596–606, 2013.
- [12] Eugene Garfield. Citation indexes for science. a new dimension in documentation through association of ideas. *International journal of epidemiology*, (5):1123–1127, 2006.
- [13] Simone Teufel, Jean Carletta, and Marc Moens. An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, pages 110–117. Association for Computational Linguistics, 1999.
- [14] Hidetsugu Nanba and Manabu Okumura. Towards multi-paper summarization using reference information. In *IJCAI*, pages 926–931, 1999.
- [15] Adam Meyers. Contrasting and corroborating citations in journal articles. In *RANLP*, pages 460–466, 2013.
- [16] Chen-Tse Tsai, Gourab Kundu, and Dan Roth. Concept-based analysis of scientific literature. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1733–1738. ACM, 2013.
- [17] Mohammad Abdullatif, Yun Sing Koh, and Gillian Dobbie. Unsupervised semantic and syntactic based classification of scientific citations. In *Big Data Analytics and Knowledge Discovery - 17th International Conference, DaWaK 2015, Valencia, Spain, September 1-4, 2015, Proceedings*, pages 28–39, 2015.
- [18] Mohammad Abdullatif, Yun Sing Koh, Gillian Dobbie, and Shafiq Alam. Verb selection using semantic role labeling for citation classification. In *Proceedings of the 2013 workshop on Computational scientometrics: theory & applications*, pages 25–30. ACM, 2013.
- [19] Jean Carletta. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22:249–254, 1996.
- [20] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL (1)*, pages 1555–1565, 2014.
- [21] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.
- [22] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 2493–2537, 2011.
- [23] Ahmad Parsian and Nader Nematollahi. Estimation of scale parameter under entropy loss function. *Journal of Statistical Planning and Inference*, 52:77–91, 1996.
- [24] Dan Jurafsky. *Speech & language processing*. Pearson Education India, 2000.