Neurocomputing 335 (2019) 195–205

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multi-task learning model based on recurrent convolutional neural networks for citation sentiment and purpose classification

Abdallah Yousif[a], Zhendong Niu[a,b,*], James Chambua[a], Zahid Younas Khan[a]

[a] School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[b] School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 11 May 2018
Revised 6 October 2018
Accepted 11 January 2019
Available online 17 January 2019

Communicated by Fenza Giuseppe

Keywords:
Citation sentiment
Convolutional
Citation purpose
Multitask learning

ABSTRACT

Automated citation analysis is a method of identifying sentiment and purpose of citations in the citing works. Most of the existing approaches use machine learning techniques to boost the performance of citation sentiment classification (CSC) and citation purpose classification (CPC), which are the main tasks of automated citation analysis. However, such approaches address CPC and CSC by learning them separately, which often suffer from inadequate training data and time-consuming for feature engineering. To alleviate these problems, we propose a multitask learning model based on convolutional and recurrent neural networks. The proposed model benefits from jointly learning CSC and CPC by modeling the citation context with task-specific information and shared layers for citation sentiment and purpose classification. The network architecture of the proposed model is useful to represent the citation context and extracts the features automatically. We conduct experiments on two public datasets to evaluate the performance of the proposed model using standard metrics such as precision, recall, and F-score. The results of CSC and CPC tasks show improvements relative to classical machine learning algorithms such as SVM and NB as well as single-task deep learning models.

1. Introduction

In general, an author's sentiment and the purposes of citing a particular paper can be determined by using automated citation analysis through analyzing the citation context of the citing papers. Citation context refers to the sentence or several sentences which are carrying the author's sentiment and purpose for citing a particular paper [1]. The classification of citation context could provide more information about author's purpose and sentiment behind quoting from the cited works [2]. In addition, existing bibliometric measures for estimating the impact of the papers provide quantitative indicators of how good a published paper is [3], this assessment often does not reflect the quality of the work presented in the article. Therefore, we believe this kind of analysis could be useful if the qualitative aspects would be added to the traditional bibliometric measures such as an Impact Factor of journals using Automated Citation Analysis (ACA).

In the literature, ACA generally refers to the tasks of citation purpose and sentiment classification. Citation sentiment is used to assign polarity (positive, negative or neutral) to the citations by analyzing the citation context of the citing papers [4]. On the other hand, citation purpose is a detailed analysis of citations and can categorize the citations into different purposes or functions [5]. The purpose and sentiment of citations could be useful for citation summarization and automated survey generation [6]. Throughout this paper we use the terms 'purpose' and 'function' interchangeably.

ACA tasks are mainly solved by using machine learning techniques explicitly adapting supervised classifiers such as SVM, Naïve Bayes (NB), logistic regression and decision trees. The supervised approaches often employ features such as part of speech (POS) tags, n-grams, negation, dependency relation, semantic patterns, cue words, and author information [6–12]. Although existing approaches have reported good results, ACA tasks remain a challenge due to insufficient training datasets and the time-consuming feature engineering. Also, due to the use of different citation schemes, it is difficult to compare the reported results.

Recently, deep learning techniques have been applied for useful representation of citations by using word embedding techniques with the aim of improving the performance of ACA systems. Munkhdalai et al. [13] proposed a model based on a variation of the recurrent neural network (RNN) namely long short-term memory (LSTM) network and attention mechanism for ACA and proposed a binary scheme with four categories for citation purpose

* Corresponding author.
E-mail addresses: abdosif33@hotmail.com (A. Yousif), zniu@bit.edu.cn (Z. Niu), jchambua@bit.edu.cn (J. Chambua), zyounask@gmail.com (Z.Y. Khan).

https://doi.org/10.1016/j.neucom.2019.01.021
0925-2312/

and sentiment classification. Their results demonstrate the effectiveness of the use of deep learning in ACA. Even though their sentiment classes in the proposed citation scheme seem to be suitable for CSC but the citation purposes/functions are not very useful for improving citation analysis methods than earlier schemes. Lauscher et al. [14] proposed to use convolution neural network with word embedding to enhance the performance of ACA. They used the citation scheme from [3] which has six citation functions for CPC and three labels for CSC. Their proposed model outperforms classical machine learning methods in terms of F-score. Despite the fact that existing approaches manage to accomplish a good classification performance, they still address CPC and CSC by learning them independently. As a result, they often suffer from insufficient training data and time-consuming for feature engineering.

In this paper, we propose a multitask learning (MTL) model based on recurrent convolutional neural networks that benefit from jointly learning CSC and CPC. Due to the correlation between our tasks the proposed model jointly learns CSC and CPC while leveraging the shared information to improve the classification performance. Our proposed model is inspired by the earlier work in multitask learning [15] and based on the relationship between CSC and CPC as presented in [16]. We exploit the relation between CSC and CPC settings to formulate them as MTL problem. MTL has shown great success in various domains and its advantages have been validated [17–19]. The proposed MTL based on deep learning architecture can model the citations with task-specific and shared layers to detect author's sentiment and purpose in citing works. To the best of our knowledge, this paper is the first attempt to introduce the application of multitask learning based on a deep learning framework for citation sentiment and purpose classification tasks. The proposed model uses the classification scheme proposed in [3] which has three and six categories for citation sentiment and purpose respectively. In addition, we have also applied another classification scheme proposed by Dong and Schäfer [16]. The experimental results on two public datasets show that our proposed model improves the performance of the ACA tasks and outperforms the baseline models.

The rest of the paper is organized as follows. Section 2 presents the related work. Section 3 describes the proposed approach. Experiments and Evaluation are presented in Section 4. Section 5 concludes the paper and recommend future works.

2. Related work

In this section, we briefly introduce existing approaches for identifying the sentiments and purposes of the citations in scientific publications.

2.1. Citation sentiment classification

Many studies have been conducted to handle CSC task, Athar [4] extracted citation sentences from ACL (Association for Computational Linguistics) Anthology corpus [20] and manually built labeled citation dataset. SVM and NB with syntactic features were used, and the combination of dependency relations and negation features achieves best classification results. Likewise, Athar and Teufel [21] expanded the citation context by identifying extra words as implicit citations to improve CSC. In their work, the use of implicit and explicit citations achieve good results compared with using explicit citation sentence only. Moreover, experiments on citation classification were carried out by Radev et al. [6] using a supervised sequence labeling technique for finding out the citation context of a given citation with the aim of classifying citations as positive or negative. Since the reference's style is different in the journals and can affect feature extraction, the authors applied a regular expression to clean the citation context and extract four citation sentences as a window. SVM was trained with self-citations, negation, dependency relations as features and achieved F-score of 71%. Furthermore, Parthasarathy and Tomar [22] used a sentence parser to extract citation sentences from random papers selected from Google Scholar and identified adjectives that could be either positive or negative. They proposed to use machine learning classifiers and concluded that if a sentence does not have an adjective, it can be either unknown or neutral. Their experiments were conducted using small dataset which can affect the classification performance. Sula and Miller [23] designed a tool to extract citation sentences and detect sentiment in the papers from humanities domain. Their tool is based on the use of NB algorithm with n-grams model as features for identifying citation polarity. They suggested that increasing the number of labeled sentences can achieve better classification results.

Recently, Kim and Thoma [24] proposed an approach to detect author's sentiment in the biomedical domain. Their method first classifies the papers as citing or cited documents, and then citation sentences are identified from the citing papers. The proposed model employed SVM with n-grams as a feature set. Xu et al. [25] developed a rule-based method to extract citations and four thousand citations were manually annotated to build their dataset. SVM with n-grams and sentiment lexicon features were used to classify the citations into either positive or negative class. It was noted that, a combination of the features achieved a better F-score than using only individual features. In [26], the authors developed annotation methodology for labeling citation sentences to be positive, negative or neutral to address the citation sentiment classification. Keywords and semantic patterns were used to discriminate the citation sentiment classes. They used SVM and managed to achieve good results. Ma et al. [17] exploited author's reputation information and proposed to use features such as unigram, polarity distribution, author information, and p-index. They reported best classification performance through the combination of author information and p-index feature.

Lexicon based approaches are useful to calculate the orientation for a document and detect the sentiment and have been applied in many sentiment analysis domains [27–29]. In CSC, Sendhilkumar et al. [30] proposed a sentiment lexicon approach to identify citation sentiment with the aim of calculating article quality. They extracted citation context as sentences and then applied part of speech (POS) tags to determine adjectives. Their method scores each adjective as positive or negative and then aggregates all scores to get the final sentiment of the citations. Similarly, Goodarzi et al. [31] proposed a framework based on the use of SentiWordNet [32] with other lexicons for citation sentiment determination. Their results showed the superiority of SentiWordNet compared with the other dictionaries in the citation sentiment classification task.

More recently, in affective computing and sentiment analysis, many techniques including deep learning approaches and distribution word representation have been used for best text representation and capturing the sentiments about public events, political activities, and product reviews [33–39]. In the citations domain, few deep learning models were used such as CNN and LSTM to address CSC task, and their results demonstrated that the use of deep learning approaches is better to capture the local and sequential features automatically as well as improving the classification performance [14,40].

Although the approaches mentioned above are interesting, most of them are domain dependent and suffer from the time-consuming feature engineering process. Furthermore, it is difficult to experiment and compare them because some of their datasets are unavailable and they utilized different schemes. In addition, to the best of our knowledge, these methods ignore the shared information between CSC and CPC which are important to improve the performance of both tasks.

2.2. Citation purpose classification

Many approaches and schemes have been proposed to deal with the citation purpose classification. Regarding the citation schemes, Garfield [41] introduced a citation scheme with fifteen functions to answer the question why authors are quoting from previous papers. Moravcsik and Murugesan [42] proposed another citation scheme which contained four dimensions comprising of conceptual, juxtaposition, perfunctory and confirmative. According to the citation role in the author's argument, Teufel et al. [43] proposed a scheme with seven functions which included background, other, own, aim, textual, contrast, and basis. Nanba and Okumura [44] proposed a simple scheme which includes three functions namely: Comparison, Basis, and Other. Then a rule-based approach was proposed with a set of cue word features to determine the function of the citations for scientific paper summarization. Teufel and Siddharthan [45] adopted 12 functions as a citation classification scheme and reduced them to four categories namely: weakness, contrast, positive and neutral. They further proposed to use support vector machine (SVM) for citation classification. In their scheme, they considered leaf nodes for citation function and the top nodes for citation sentiment. Hernandez-Alvarez et al. [46] proposed a citation scheme with six functions and utilized machine learning techniques to detect author's sentiment and purpose for citing a paper. However, most of the existing works are limited to the domain of computer science, and the functions are overlapped and complex. Moreover, it is also difficult to use these schemes for labeling a sentence.

In automated CPC, different techniques have been proposed including rule-based and machine learning techniques. Nanba and Okumura [47] used a rule-based approach with the aim of reducing the functions into three functions to ensure easy annotation of the citations. Moreover, Teufel and Siddharthan [45] randomly selected 116 articles and mapped the existing twelve citation functions into four categories to improve the annotation of the citations. Nakagawa et al. [48] employed a supervised model based on the conditional random fields to detect citation functions in the Japanese language. Moreover, Dong and Schäfer [16] proposed to use NB with features such as negation, cue words, POS-tag, location, and popularity. Their methods categorized the citations into four functions. Jochim and Schutze [49] used a maximum entropy classifier and a window size of three sentences, which captured the whole part of the cited work according to the citing work. Meyers [50] proposed a hybrid method by using discourse as a tree model and analyzed POS tags to find out citation relations regarding contrast and corroboration functions. Furthermore, Tsai et al. [51] classified citations into applications and techniques by using an unsupervised bootstrapping algorithm. To automatically generate a citation scheme, Abdullatif et al. [52] proposed a rule-based approach based on semantic role labeling to select the relevant verb that represents the citation sentence. On the other hand, ontologies are essential for knowledge representation, in this context, Di Iorio et al. [53] developed software called CiTaLo using ontology learning and natural language processing techniques for detecting citation polarity. Furthermore, they employed CiTaLO tool, CiTo2wordnet and CiTo function to retrieve citation functions. They used positive, negative, and neutral functions or classes (e.g., agree, cites, cites as authority, cites a data source and cites as evidence).

Recently, Abdullatif et al. [54] used clustering techniques and proposed models for computing the similarity between citation sentences. In their model, each group of similar sentences was considered as a citation function. Although previous approaches address CPC effectively, the solutions are based on single task learning by training separate models for citation purpose and citation sentiment, unlike our proposed multi-task learning approach.

3. Proposed multitask-RCNN model

In this paper, we propose a model to address citation sentiment and citation purpose classification by encoding citation sentiment and purpose information in a word embedding representation, which consequently used by neural network architecture with MTL to improve the classification performance. Fig. 1 illustrates the architecture of the representation learning network (Multitask-RCNN), which encodes citation sentiment and purpose information in word embedding. The architecture comprises a combination of two neural networks, which are CNN and BiLSTM for extracting n-gram features and capture long-term dependencies of the input citation sentence respectively. CNN has been extensively used for syntactic and semantic representations of the text in various NLP tasks and has been verified to attain better performance than classical NLP methods.

Generally, in CNN, the kernel size is much smaller than the input size. As a result, the output is only interacting with a narrow window of the input and usually highlights the local lexical connections of the n-gram. RNN is well designed for sequence modeling and its variation (LSTM and Bi-LSTM) can efficiently keep the valuable information from long-range dependency but may fail to capture the local n-gram context. Essentially, CNN and LSTM have their advantages and disadvantages, and it has been found that using a vector from either CNN or LSTM to encode an entire sequence is not sufficient to capture all the important information. Thus, we propose RCNN architecture, which combines CNN and BiLSTM by exploiting the advantages from both algorithms. We have noted that our architecture as depicted in Fig. 1 employs LSTM before CNN comparing to [55–57] in order to capture the global relationships, especially word order information, among the terms in the sentence. Then we use CNN to capture the local relationships (or latent neighbor relationships) between terms in the sentences.

In the proposed approach, we consider a citation context with four sentences. The four sentences are concatenated in order to capture more opinionated words which are useful in detecting the sentiment and purpose of citations. The proposed architecture converts the input sentences into word vectors by using word embedding technique. Since the length of the citation sentences is varying, we pad the sentence vectors with zeros to get the same length. Then, BiLSTM network is used to capture long-term dependencies followed by CNN for extracting n-gram features. The final representation is fed to the multitask output classification, which is a fully connected layer with multitask outputs to produce the scores for citation sentiment and purpose classification. Algorithm 1 summarizes the major components of the proposed model and the detailed description of each component is presented in the following subsections.

3.1. Recurrent neural networks (RNNs)

RNNs are neural networks suitable for sequence modeling and are the basic block in our network architecture. We use a variant of RNNs which is bidirectional long short-term memory (BiLSTM) network [57] because, in most of the cases, single direction LSTM is insufficient and suffers in utilizing the contextual information from the future words. The BiLSTM tend to learn long-term dependencies better than vanilla RNNs. Therefore, the Bi-LSTM utilizes both previous and next contexts by processing the sequence on both forward and backward directions. Furthermore, this method can solve the gradient vanishing problem [58]. Recently Shuai

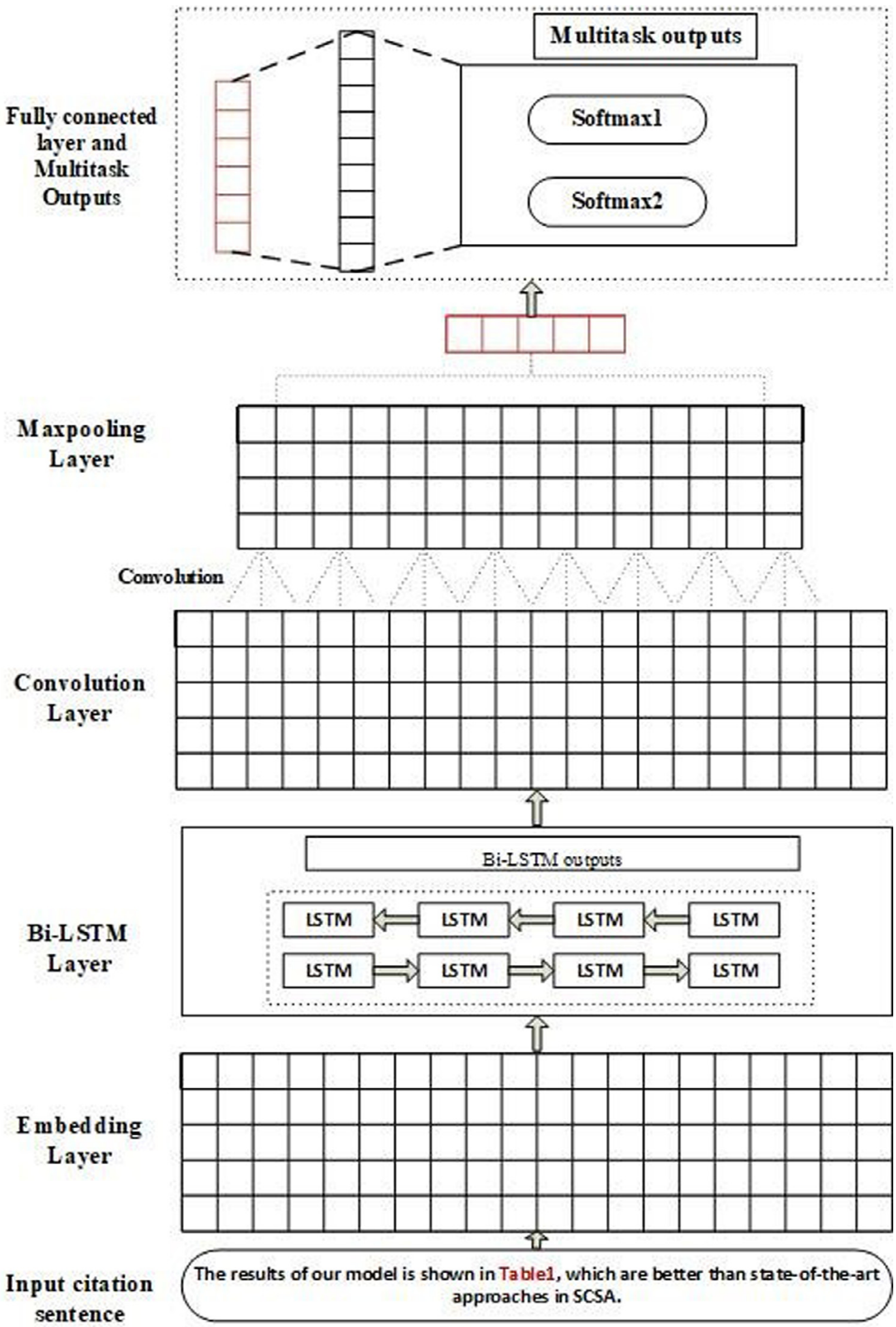


Fig. 1. Multitask-RCNN architecture for automated citation analysis.

et al. [58] proposed IndRNN method which can effectively solve this problem for very deep networks. However, in this initial attempt, given the size of our datasets and the depth of our network architecture, we believe that simple BiLSTM is sufficient for our task. The original citation sentence S is a sequence of words: $w_1, w_2, \ldots w_s$ where each word is drawn from a finite-size vocabulary V with size |V| are represented with word embeddings via a lookup table operation:$v_i = LT_W(w_i)$, where $W \in \mathcal{R}^{\mathrm{d} \times |V|}$ is the word embedding matrix and d is the embedding dimension that represents the corresponding word.

The input to BiLSTM is a sequence of vectors $= s_1 s_2 \ldots\ s_l$, and the output is a sequence of vectors $H = h_1 h_2 \ldots\ h_l$. At each time step t, the input gate i_t, forget gate f_t, output gate o_t, cell state c_t and one output vector h_t are calculated using the following equations.

$$f_t = \sigma\left(W_f \cdot s_t + \mathrm{U}_f \cdot h_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_i \cdot s_t + \mathrm{U}_i \cdot h_{t-1} + b_i\right) \tag{2}$$

$$\tilde{c}_t = \tanh(W_c \cdot s_t + \mathrm{U}_c \cdot h_{t-1} + b_c \tag{3}$$

$$o_t = \sigma\left(W_o \cdot s_t + \mathrm{U}_o \cdot h_{t-1} + b_o\right) \tag{4}$$

Algorithm 1 Multitask-RCNN algorithm.

Input: the input consists of a sequence of words: $w_1, w_2, \ldots w_s$, where w_i is drawn from a finite-sized vocabulary V
1. Represent w_i with its corresponding word embedding $v_i = LT_W(w_i)$ and $S = s_1 s_2 \ldots s_l$ as the input sentence embedding matrix with dimension $R^{d \times |s|}$
2. Apply BiLSTM on the input matrix S to get output vectors $h_t \in H = (h_1, h_2, \ldots h_l)$
3. Apply CNN by using a set of filters $k_{i:n}$ on H to produce the feature maps C, where $C_{it} = k_i^T.h_t$
4. Apply the rectified linear unit to process C with the outputs X defined as $X_{it} = \max(0, C_{it})$
5. Apply max-pooling to process X and get $\breve{X}_s \in R^n$ where $\breve{X}_s[i] = \max(0, X[i, :])$.
6. Apply softmax $(\breve{X}_s)$ and produce $y_i'^t = f(x_i^t.W^t + b^t)$ where f is a softmax function of x_i^t parameterized by a weight matrix W^t and bais term b^t.

Output: return y'^t where t is the task.

$$c_t = i_t \otimes \tilde{c}_t + f_t{}^{\circ} c_{t-1}) \tag{5}$$

$$h_t = o_t \otimes \tanh(c_t)) \tag{6}$$

Where s_t and h_t are the input to Bi-LSTM and output vectors at time t, respectively. W_f, W_i, W_o, U_i, U_f, and U_o are weight matrices and b_i, b_f, b_c, and b_o are bias vectors. The symbol $\otimes$ denotes element-wise multiplication and σ represents the sigmoid function. The input gate controls how much information desires to flow into the memory cell while the forget gate picks what information is essential to be removed from the memory cell. The output gate produces the hidden state for the current input word. The output representation vector h_l is then given to a CNN to capture most important features.

3.2. Convolutional neural networks (CNN)

The one-dimensional convolutional involves a filter vector sliding over a sequence and detecting features at different positions. Given the output of Bi-LSTM network $H \in R^{d \times l}$, the convolution uses a set of n filter vectors $k_i \in R^{ml}$, with sliding window size m to process it by a linear convolution operation. Formally, let $H_{i:i+j}$ refers to the concatenation of $h_i, h_{i+1}, \ldots, h_{i+j}$. The linear convolution operation takes the dot product of k_i with each $m-$gram in the sentence S as shown in Eq. (7).

$$C_{it} = k_i^n.H_{n-m+1:n} \tag{7}$$

Where n is the number of filters, $C \in R^{n \times d}$, we use a rectified linear unit as a non-linear function. Pooling operation is used to extract robust features from the output of the convolution operation. Max-pooling is applied for each filter to capture the most significant information. The max value is extracted from each row of C, which generates the final representation vector $\breve{X}_s \in R^n$ for the input sentence S.

3.3. Multitask output classification

The idea in deep learning is to have several neural networks with different sets of parameters to address different tasks. However, as suggested and carried out by many researchers, an effective approach to handle several tasks is to use multitask learning approach by sharing some of the network architecture and its parameters. To address our tasks, one solution is to build two networks for CSC and CPC, and each network has different sets of parameters. Instead of training a separate network for each task, we build a single network with two outputs (Fig. 1) and share some of our network architecture and parameters. As a result, learning CSC and CPC jointly is more useful for citation classification than training the model for each task individually. As depicted in the top part of Fig. 1, after getting the outputs of CNN $\breve{X}_s$, which is the entire representation (shared representation) of the input text S. we fed $\breve{X}_s$ to the fully connected layer which has two softmax functions to produce the class probabilities of each task. The following steps are the description of model training as well as the prediction of the outputs.

Suppose we have a total of T task and the training data for the tth task are denoted as

$(x_i^t, y_i^t)_{1=1}^{D_t}$

where $t = 1, \ldots, T,\ i = 1, \ldots,\ D_t$, D_t is the number of samples of the t-th task, with x_i^t and y_i^t being the feature vectors and actual class values, respectively. We applied a softmax to obtain the prediction $y_i'^t$ for each task as follows:

$$y_i'^t = f\left(x_i^t.W^t + b^t\right) \tag{8}$$

where f is the softmax function of x_i^t parameterized by a weight matrix W^t and bias term b^t, the aim of MTL is to minimize the loss as shown in Eq. (9).

$$argmin \sum_{i=1}^{D} L\left(y_i^t, y_i'^t\right) \tag{9}$$

where L is the loss function. There are several loss functions for the classification task. Here, we used a commonly used cross entropy function to calculate the loss of the tth task which can be formally defined as:

$$Loss\,of\,Task\,t = -\sum_{i=1}^{D} y_i^t.\log\left(y_i'^t\right) \tag{10}$$

Furthermore, we apply dropout strategy [59] to avoid overfitting problem which usually affects the classification performance. In addition, we customize the loss function Eq. (10) by adding a regularization term on weight vectors, which gives us an improved loss function for t-th task as defined in Eq. (11).

$$Loss\,of\,Task\,t = -\sum_{i=1}^{D} y_i^t.\log\left(y_i'^t\right) + \beta \left\|W^t\right\|_2^2. \tag{11}$$

where $\|.\|$ denotes the L_2 norm and β represents a penalty constant. The training details of our model are further introduced in Section 4.1.3.

4. Experiments and evaluation

In this section, we provide the implementation details and analysis of the proposed Multitask-RCNN model, and the comparison of experimental results. Section 4.1 shows how we initialized the hyperparameters of our experiments as well as presents the datasets that we used and the baseline models. Section 4.2 presents the macro-average precision, recall, and F-score of the proposed model from the benchmark datasets and the comparison with the baseline models. In addition, a brief discussion about the impact of hyper-parameters are presented in Section 4.3.

4.1. Experimental settings

4.1.1. Dataset

The proposed multitask-RCNN architecture for citation classification is evaluated using two publicly available datasets. The first dataset (Dataset1) from [6,60] has 3568 citation context examples. Each example contains four sentences: the sentence quoting from a

Table 1
The statistics of the datasets.

Layer	Parameter name	Parameter value
Lookup	Word embedding dim	50, 100, 200,300
CNN	Window size	3,4,5
	Number of filters	50,100, 150, 200
LSTM	Hidden units	50, 100, 150, 200
Dropout	Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8
	epochs	10, 20, 30, 50
	Batch size	16, 32, 64, 128
	Bernoulli probability	0.2, 0.3, 0.4, 0.5

Table 2
Hyper-parameters settings.

Classification task	Dataset1		Dataset2	
	Class	*Ratio (%)*	*Class*	*Ratio (%)*
Citation sentiment	Positive	32.6	Positive	10.75
	Negative	12.4	Negative	3.22
Citation purpose	Neutral	55.0	Neutral	86.03
	Criticizing	16.3	Idea	7.18
	Comparison	8.1	Basis	23.81
	Use	18.0	GRelated	42.48
	Substantiating	8.0	SRelated	20.81
	Basis	5.3	MRelated	1.75
	Neutral	44.3	Compare	3.97

given cited paper, one previous sentence, and two subsequent sentences. These contexts have been annotated for both citation sentiment and citation purpose. Citation sentiment has three categories: positive, negative, or neutral. Whereas the citation purpose was annotated as a scheme of six functions: criticism, comparison, use, substantiation, basis, and neutral. The second dataset (Dataset2) [16] is built based on randomly extracted citation sentence from the ACL corpus,[1] which has 1768 instances. Dataset2 is annotated for both citation sentiment and purpose classification. Citation purpose was labeled as a scheme with six functions: idea, basis, Grelated, Srelated, Mrelated, and compare. Furthermore, citation sentiment is annotated as positive, negative or neutral. The distributions of instances over the different categories for both classification tasks are shown in Table 1.

4.1.2. Baseline models and metrics

We compare our model with the following baseline approaches for citation sentiment and purpose classification:

- NB with syntactic features [16]
- SVM with rich features [3]
- SVM with TF-IDF [14]
- SVM with word embedding [14]. In this model, the input is features generated using word2vec method.
- CNN with word embedding [14].
- Since the proposed model consists of CNN and Bi-LSTM, we also reported the performance of Single task-BiLSTM [40], Single-task-RCNN, Multitask-CNN, Multitask-BiLSTM.

We used same evaluation metrics as presented in [14], which are average Macro-precision (P), recall (R), and F-score (F) for sentiment and purpose categories.

4.1.3. Hyper-parameters

In all experiments, we followed the settings used in [14], that is, applying 10-fold cross with grid search method to select best hyper-parameters from the options presented in Table 2. We used Adam optimizer to train our model and the losses were calculated by using categorical cross entropy. To train our model, we adopted the proposed procedure in [61] which works as follows: before picking the batch which is determined by grid search, we perform a Bernoulli trial with probability p and $1 - p$ to pick a batch for citation sentiment and purpose classification respectively. The error is back propagated to the embedding layer and the weights of the network are shared and therefore affected by both tasks.

In the case of baseline models (SVM, and NB), each dataset was split by 80%, 10%, and 10% for training, development, and test sets respectively. We implemented the baseline models using Dataset2. For the SVM model, we performed a grid search using development set to select the best hyperparameters for evaluation. The final SVM model was learned on both the training and development sets and tested on the test set based on the best-obtained hyperparameters. The CNN models were trained only on the training set whereas SVM model was built on both training and development sets. We used the development set to evaluate the other neural network models (CNN, BiLSTM, Multitask-CNN, and Multitask-BiLSTM) and the final performance of the models was evaluated on the test sets. For the experiments, we used python as a programming language of our choice. Furthermore, to implement the proposed model and other deep learning algorithms we used Keras library[2] to implement the proposed model. Additionally, Scikit-learn[3] was used to implement the baseline models.

Table 3
Comparison results of citation sentiment classification.

Method	Dataset 1			Dataset 2		
	P	*R*	*F*	*P*	*R*	*F*
NB with syntactic features [16]	69.0	62.5	64.4	–	–	72.00
SVM with features [9]	67.1	70.6	68.8	70.08	56.66	54.48
SVM with TF-IDF [14]	77.9	76.3	77.1	75.13	61.71	59.53
SVM with Embedding [14]	81.3	75.4	77.3	79.21	77.57	77.90
CNN with Embedding [14]	82	75.9	78.8	85.21	81.02	82.86
Single task-LSTM	80.8	74.3	77.4	83.26	82.14	82.12
Single task-BiLSTM [40]	80.4	77.56	79.1	84.76	82.41	82.97
Single task-RCNN	85.17	82.45	83.69	86.14	84.47	85.27
Multitask-CNN	89.31	86.55	87.91	85.71	84.75	85.23
Multitask-BiLSTM	86.48	85.99	85.19	84.1	83.62	83.85
Multitask-RCNN	92.10	84.87	88.34	88.13	85.18	86.63

Table 4
Comparison results of citation purpose classification.

Method	Dataset 1			Dataset 2		
	P	*R*	*F*	*P*	*R*	*F*
NB with syntactic features [16]	65.02	58.5	60.4	–	–	72.0
SVM with features [9]	54.9	62.5	58.4	62.1	59.57	60.0
SVM with TF-IDF [14]	74.3	70.9	72.6	64.9	62.5	62.8
SVM with Embedding [14]	86.8	64.7	74.1	79.6	67.6	73.1
CNN with Embedding [14]	80.8	68.8	74.3	84.6	71.6	76.14
Single task-LSTM	79.87	67.8	73.21	82.3	69.2	74.1
Single task-BiLSTM [40]	77.22	73.11	75.11	78.02	73.95	75.91
Single task-RCNN	79.12	76.87	77.98	80.12	77.87	78.65
Multitask-CNN	83.92	80.39	82.12	80.72	77.19	78.92
Multitask-BiLSTM	87.1	75.63	80.96	85.90	74.63	79.76
Multitask- RCNN	85.11	84.13	84.62	84.04	82.15	83.08

4.2. Results and discussion

The comparison results of the proposed model against other baseline models are presented in Table 3 (citation sentiment classification) and Table 4 (citation purpose classification). The results generally indicate that for both sentiment and purpose classi-

[1] http://aclweb.org/anthology.

[2] https://keras.io/.

[3] http://scikit-learn.org/stable/.

fication tasks, our proposed approach outperforms the baseline methods. We grouped the methods into three categories: traditional machine learning classifiers with various features, single task neural network classifiers, and multitask neural models.

In the first category, the results of top 5 methods have been previously reported in [14] for both tasks, i.e., citation sentiment and purpose classification. In this category, as can be seen from Table 3 the first four approaches are classical classifiers with handcraft features. As presented in Table 3, we observed that SVM with embedding features performs better than NB and SVM with handcraft features such as syntactic and TF-IDF features on Dataset1 and Dataset2 and this is due to the benefits of using word representation techniques. Similarly, as shown in Table 4, SVM classifier with embedding features achieves better performance on the citation purpose classification.

In the second category, single task learning models (CNN, LSTM, BiLSTM, and RCNN) use word embeddings for citation classification. As can be seen in Tables 3 and 4, single task RCNN, which is a combination of CNN and LSTM, achieved the best performance compared with individual CNN, LSTM, and BiLSTM in Dataset1 and Dataset2. These results indicate the advantage of combining both algorithms. Moreover, single task learning models achieve better performance compared with traditional classifiers and this is due to the effectiveness of deep learning approaches in improving the classification performance.

The third category includes the Multitask-CNN, Multitask-BiLSTM, and Multitask-RCNN. These models use MTL and achieve better performance than the baselines in other categories which demonstrate that MTL is a useful technique for improving citation sentiment and purpose classification. Moreover, we noted that the proposed multitask-RCNN approach for both tasks outperforms classical classifiers and single task neural network models. These findings validate the impact of MTL in improving the classification performance.

In detail, as depicted in Table 3, the performance of the approaches on each of the two datasets is reported based on macro-average precision, recall, and F-score values. It is observed that Multitask-RCNN obtains best results for the citation sentiment classification task on Dataset1. The model improves results by 4.65% in terms of F-score (88.34%) comparing with the single task-RCNN F- score (83.69%). We further noted that comparing with the traditional classifiers the proposed model is superior in its performance and reported an increase in F-score by more than 11%. Furthermore, for Dataset2, the proposed Multitask-RCNN model outweighs classical classifiers and single task neural models in all metrics. Multitask-RCNN achieved an F-score of 86.63% which is close to more than 1% increase over the single task RCNN. However, this marginal increase can be justified by two reasons. First, the size of the Dataset2 is smaller comparing to Dataset1. Second, the classification on Dataset2 was based on one citation sentence as citation context size whereas in Dataset1 four citation sentences were used. The size of the citation context has been shown to affect the classification performance. We discuss the details of the impact of citation context later in this section.

As depicted in Table 4, on citation purpose classification our model beats the baseline methods as well. On Dataset1, the proposed model achieves an F-score of 84.62%, which is equivalent to the performance increase of 6% over the F-score recorded (77.98%) for single task neural models. In addition, for Dataset2 the proposed model also achieved a growth of 4% for F-score. Moreover, comparing the classical classifiers with each other, on both tasks, SVM with word embedding features leads to better results than NB with syntactic features. Note that, the results of SVM on Dataset2 for both tasks were not good as the proposed features in [16] because we have only experimented by using n-grams features on this dataset. On the other hand, SVM with embedding features on Dataset2 obtained better results compared with NB. This is due to the advantage of using word embedding techniques.

Table 5
Impact of citation context size on the citation sentiment classification.

Method	Context size	*P*	*R*	*F*
SVM with Embedding [14]	Cited sentence	83.2	72.1	75.3
	Four sentences	84.1	75.6	79.6
CNN [14]	Cited sentence	81.8	76.1	78.8
	Four sentences	85.8	78.7	82.1
Multitask-RCNN	Cited sentence	88.83	86.83	87.82
	Four sentences	92.10	84.87	88.34

Table 6
Impact of citation context size on the citation purpose classification.

Method	Context size	*P*	*R*	*F*
SVM with Embedding [14]	Cited sentence	81.7	66.2	73.1
	Four sentences	84.8	69.2	76.2
CNN [14]	Cited sentence	80.8	68.8	74.1
	Four sentences	85.2	73.3	78.9
Multitask-RCNN	Cited sentence	84.24	82.35	83.28
	Four sentences	85.11	84.13	84.62

Table 7
Impact of dropout strategy.

Classification task	Option	*P*	*R*	*F*
Citation sentiment	Non-dropout	88.83	86.83	87.82
	Dropout	92.10	84.87	88.34
Citation purpose	Non-dropout	83.81	82.63	83.22
	Dropout	85.11	84.13	84.62

Moreover, we assessed the impact of citation context size on the classification performance of our model by using Dataset1 which has one sentence and four sentences as a citation context. After the series of experiments, the results generally indicated that the performance increases with an increase of the citation context size as depicted in Tables 5 and 6. These results are consistent with the results reported in the previous works [14].

Furthermore, Figs. 2 and 3 show the effectiveness of the proposed approach compared with same deep learning architecture as a single task learning for citation sentiment and purpose classification. We further noted that Multi-RCNN is better in terms of precision, recall, and F-score than single task RCNN.

4.3. Impact of hyper-parameters

We conducted a number of experiments using the dropout strategy in embedding and fully connected layers to alleviate the overfitting problem. Table 7 shows the impact of dropout strategy on the experimental results of the proposed model. We noted that using dropout strategy improves the classification performance by 1.4% in F-score. Moreover, Fig. 4 shows the performance of the proposed model on both classification tasks. As can be seen, the results indicate that better F-score is attained when the dropout value of embedding layer is close to 0.7 for CSC and CPC. In addition, applying the dropout after fully connected layer (Fig. 5) gives best results when the value of dropout probability was set to 0.3 in the citation sentiment and citation classification tasks.

In summary, previous works have reported the effectiveness of machine learning classifiers with many features, in particular, SVM in improving the classification performance of ACA tasks [3,4,10]. Moreover, they focused on training a single task learning models to identify author's sentiment and purpose separately. However, these approaches have been domain dependent and suffer from insufficient training data and time-consuming feature engineering. In this paper, we developed a multitask learning model and exploit the

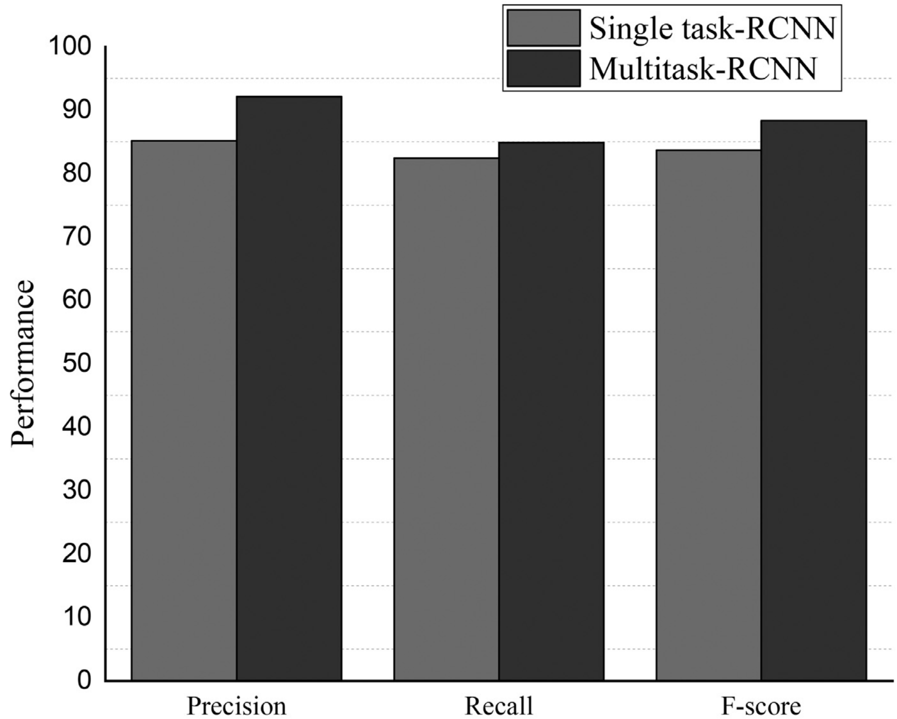


Fig. 2. Performance comparison between single task-RCNN and Multitask-RCNN on CSC.

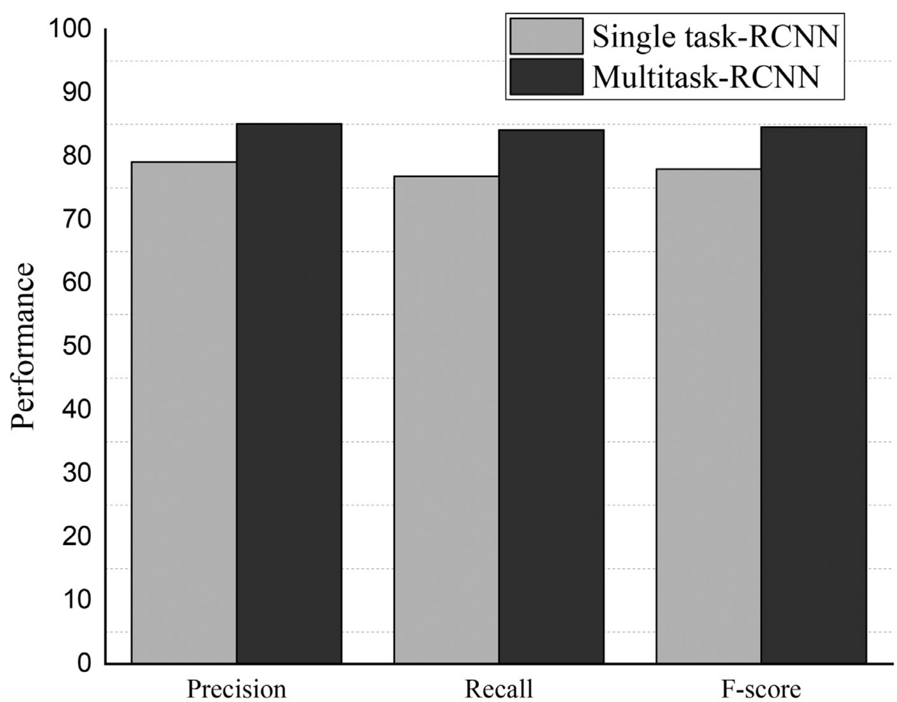


Fig. 3. Performance comparison between single task-RCNN and Multitask-RCNN on CPC.

robust of deep learning approaches to build better network architecture to represent the citations to detect the citation sentiment and purpose in citing papers. Experimental results on two public datasets show that the proposed Multitask-RCNN model effectively exploits the shared information between CSC and CPC and outperforms the baseline models. The findings of this study confirm that using MTL significantly improves the classification performance over handcraft feature based methods and single task learning approaches. Furthermore, it is evident that the benefits gained from the proposed model may address researchers' needs in developing more robust citation analysis methods to evaluate the research quality. To the best of our knowledge, this is the first work to investigate the multitask learning based on deep learning in ACA and the use of RCNN network architecture to represent the citations and capture the features automatically. Our findings suggest that the proposed approach is effective in improving bibliometric

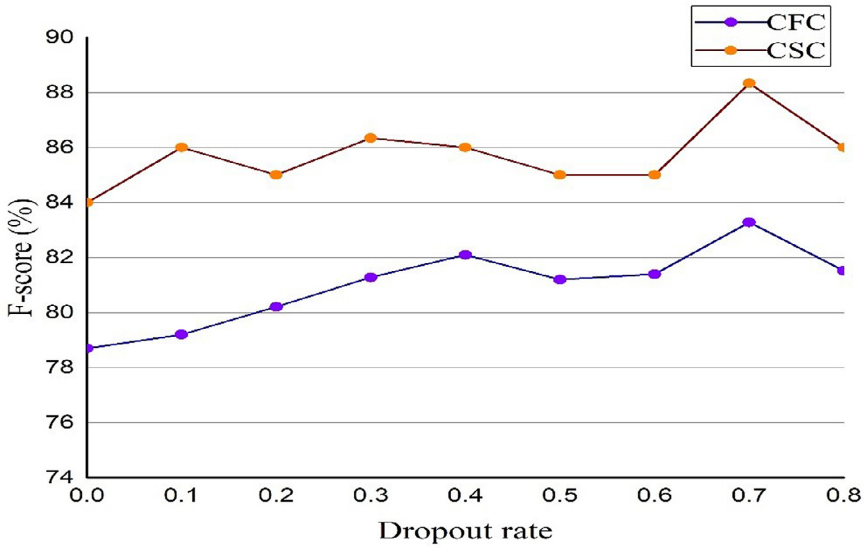


Fig. 4. Impact of different dropout rates in the word dense layer.

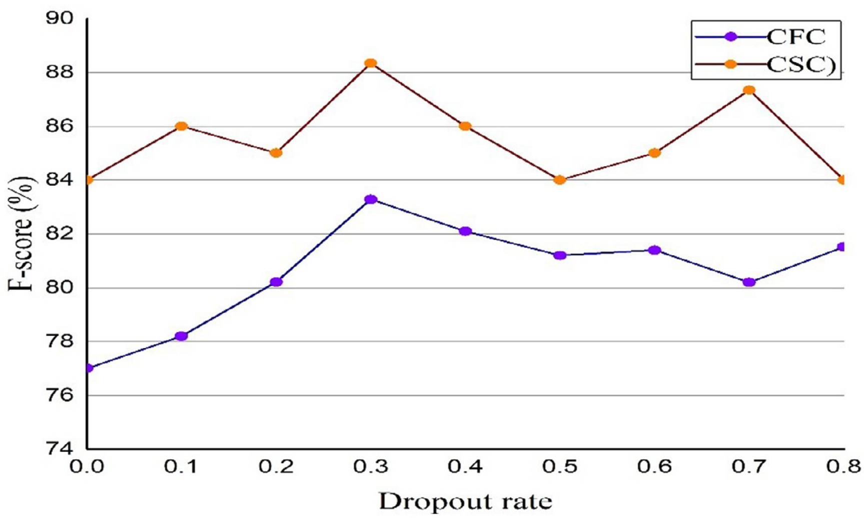


Fig. 5. Impact of different dropout rates in the embedding layer.

methods by combining the negative citations in their scheme unlike traditional citation analysis methods, which depend solely on the simple citation count to calculate the quality of researches.

5. Conclusion and future work

With the rapid growth in the number of scientific papers, it is becoming increasingly difficult to assess the quality of publications. Current bibliometric measures provide a solution to this problem but mostly are quantitative and rely on the citation count only. With analyzing citation context information, it is possible to add the qualitative aspects (sentiment and purpose) of the citations to understand its significance and helping to improve the bibliometric measures. In this paper, we proposed a multitask learning approach based on a deep learning framework to enhance the performance of automated citation analysis tasks. The proposed approach employs RCNN architecture to represent the citation context and capture the most meaningful information of the input. Using this approach, we classified the citations into three categories for citation sentiment and six categories for citation purpose. Furthermore, MTL was used to enhance the classification performance by jointly learning the citation sentiment and purpose classification tasks. We have conducted extensive experiments on public datasets and compared our approach with other ten approaches. The experimental results show that our method performs better on citation sentiment and purpose classification than the baseline approaches. We have found that citation context with the size of four sentences improves the performance of both tasks. In the future, we plan to incorporate the attention mechanism into our network architecture, since it is useful in capturing the most important information and has been proved to be convenient in text classification as general. Further, since the used datasets are small size and imbalance, semi-supervised extreme learning machine [62] might be useful technique to handle these problems and integrate additional unlabeled datasets. Finally, we also plan to investigate some other recently proposed deep learning approaches such as IndRNN and utilize recent recommendation technologies [63,64] to recommend relevent papers to the citing articles.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61370137), the National Basic Research Program of China (No. 2012CB7207002), the Ministry of Education - China Mobile Research Foundation Project No. 2015/5-9 and No. 2016/2-7 and the 111 Project of Beijing Institute of Technology.

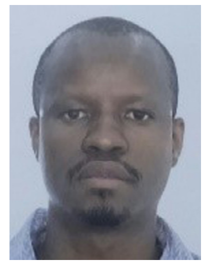
References

[1] A. Athar, S. Teufel, Context-enhanced citation sentiment detection, in: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Montreal, Canada, 2012, pp. 597–601.
[2] A. Yousif, Z. Niu, J.K. Tarus, A. Ahmad, A survey on sentiment analysis of scientific citations, Artif. Intell. Rev. (2017), doi:10.1007/s10462-017-9597-8.
[3] A. Abu-Jbara, J. Ezra, D.R. Radev, Purpose and polarity of citation: towards NLP-based bibliometrics, in: M.R. Lucy Vanderwende (Ed.), Proceedings of the North American Association for Computational Linguistics (NAACL-HLT 2013), Association for Computational Linguistics: Human Language Technologies, Atlanta, Georga, United States, 2013, pp. 596–606.
[4] A. Athar, Sentiment analysis of citations using sentence structure-based features, in: Proceedings of the ACL 2011 student session,, Association for Computational Linguistics, 2011, pp. 81–87.
[5] M. Abdullatif, Towards Unsupervised Citation Classification, ResearchSpace@ Auckland, 2015.
[6] D.R. Radev, P. Muthukrishnan, V. Qazvinian, A. Abu-Jbara, The ACL anthology network corpus, Lang. Resour. Eval. 47 (2013) 919–944.
[7] R. Radoulov, Exploring automatic citation classification, (2008).
[8] S. Agarwal, L. Choubey, H. Yu, Automatically classifying the role of citations in biomedical articles, Proceedings of American Medical Informatics Association Fall Symposium (AMIA), Washington, DC, (Citeseer2010), pp. 11–15.
[9] X. Li, Y. He, A. Meyers, R. Grishman, Towards fine-grained citation function classification, RANLP (2013) 402–407.
[10] C.A. Sula, M. Miller, Citations, contexts, and humanistic discourse: toward automatic extraction and classification, LLC 29 (2014) 452–464.
[11] M. Hernández-Alvarez, J.M. Gomez, Citation impact categorization: for scientific literature, Computational Science and Engineering (CSE), in: Proceedings of the 2015 IEEE 18th International Conference on, IEEE, 2015, pp. 307–313.
[12] Z. Ma, J. Nam, K. Weihe, Improve sentiment analysis of citations with author modelling, in: Kevin Knight (Ed.), Proceedings of the Fifth Workshop On Computational Linguistics For Literature - NAACL-HLT 2016, Association for Computational Linguistics (ACL), San Diego, California, USA, 2016, pp. 122–127.
[13] T. Munkhdalai, J. Lalor, H. Yu, Citation analysis with neural attention models, EMNLP 2016 (2016) 69.
[14] A. Lauscher, G. Glava, #353, S.P. Ponzetto, K. Eckert, Investigating convolutional networks and domain-specific embeddings for semantic classification of citations, in: Proceedings of the 6th International Workshop on Mining Scientific Publications, Toronto, ON, Canada, ACM, 2017, pp. 24–28.
[15] R. Caruana, Multitask learning, Mach. Learn. 28 (1997) 41–75.
[16] C. Dong, U. Schäfer, Ensemble-style self-training on citation classification, in: Kam-Fai Wong (Ed.), Proceedings of the 5th International Joint Conference on Natural Language Processing, Association for Computational Linguistics (ACL), Chiang Mai, Thailand, 2011, pp. 623–631.
[17] F. Jin, S. Sun, A multitask learning approach to face recognition based on neural networks, in: C. Fyfe, D. Kim, S.-Y. Lee, H. Yin (Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2008, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 24–31.
[18] C. Widmer, J. Leiva, Y. Altun, G. Rätsch, Leveraging sequence classification by taxonomy-based multitask learning, in: B. Berger (Ed.), Research in Computational Molecular Biology, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 522–534.
[19] H.G. Noushahr, S. Ahmadi, Multitask learning for text classification with deep neural networks, in: M. Bramer, M. Petridis (Eds.), Research and Development in Intelligent Systems XXXIII, Springer International Publishing, Cham, 2016, pp. 119–133.
[20] D.R. Radev, P. Muthukrishnan, V. Qazvinian, The ACL anthology network corpus, in: Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries, Association for Computational Linguistics, 2009, pp. 54–61.
[21] A. Athar, S. Teufel, Detection of implicit citations for sentiment detection, in: Proceedings of the Workshop on Detecting Structure in Scholarly Discourse, Association for Computational Linguistics, 2012, pp. 18–26.
[22] G. Parthasarathy, D. Tomar, Sentiment analyzer: analysis of journal citations from citation databases, in: 2014 5th International Conference – Confluence The Next Generation Information Technology Summit (Confluence), IEEE, 2014, pp. 923–928.
[23] C.A. Sula, M. Miller, Citations, contexts, and humanistic discourse: toward automatic extraction and classification, Lit. Linguist. Comput. 29 (3) (2014) 452–464, doi:10.1093/llc/fqu019.
[24] I.C. Kim, G.R. Thoma, Automated classification of author's sentiments in citation using machine learning techniques: a preliminary study, in: Proceedings of the 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE, 2015, pp. 1–7.
[25] J. Xu, Y. Zhang, Y. Wu, J. Wang, X. Dong, H. Xu, Citation sentiment analysis in clinical trial papers, in: AMIA Annual Symposium Proceedings, American Medical Informatics Association, 2015, p. 1334.
[26] M. Hernández-Álvarez, J. Gómez Soriano, P. Martínez-Barco, Annotated corpus for citation context analysis, Lat. Am. J. Comput. Fac. Syst. Eng. Nat. Polytech. School Quito-Ecuador 3 (2016) 35–42.
[27] X. Fu, W. Liu, Y. Xu, L. Cui, Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis, Neurocomputing 241 (2017) 18–27.
[28] S. Huang, Z. Niu, C. Shi, Automatic construction of domain-specific sentiment lexicon based on constrained label propagation, Knowl.-Based Syst. 56 (2014) 191–200.
[29] S. Park, W. Lee, I.-C. Moon, Efficient extraction of domain specific sentiment lexicon with active learning, Pattern Recognit. Lett. 56 (2015) 38–44.
[30] S. Sendhilkumar, E. Elakkiya, G. Mahalakshmi, Citation semantic based approaches to identify article quality, Proceedings of the international conference ICCSEA2013), pp. 411–420.
[31] M. Goodarzi, M.T. Mahmoudi, R. Zamani, A framework for sentiment analysis on schema-based research content via lexica analysis, Proceedings of the 2014 7th International Symposium on Telecommunications (IST), (IEEE2014), pp. 405–411.
[32] S. Baccianella, A. Esuli, F. Sebastiani, SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining, in: LREC, 2010, pp. 2200–2204.
[33] G. Rao, W. Huang, Z. Feng, Q. Cong, LSTM with sentence representations for document-level sentiment classification, Neurocomputing 308 (2018) 49–57.
[34] J. Chen, C. Zhang, Z. Niu, Identifying Helpful Online Reviews with Word Embedding Features. Knowledge Science, Engineering and Management, Springer International Publishing, Cham, 2016.
[35] H. Fu, Z. Niu, C. Zhang, J. Ma, J. Chen, Visual cortex inspired CNN model for feature construction in text analysis, Front. Comput. Neurosci. 10 (64) (2016).
[36] J. Wang, B. Peng, X. Zhang, Using a stacked residual LSTM model for sentiment intensity prediction, Neurocomputing 322 (2018) 93–101.
[37] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2016) 102–107.
[38] M. Tan, C. Dos Santos, B. Xiang, B. Zhou, Improved representation learning for question answer matching, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1, 2016, pp. 464–473. Long Papers.
[39] Y. Li, Q. Pan, T. Yang, S. Wang, J. Tang, E. Cambria, Learning word representations for sentiment analysis, Cogn. Comput. 9 (2017) 843–851.
[40] T. Munkhdalai, J. Lalor, H. Yu, Citation analysis with neural attention models, in: T.H. Cyril Grouin, Aurélie Névéol, Pierre Zweigenbaum (Eds.), Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI), Association for Computational Linguistics (ACL), Austin, TX, 2016, p. 69.
[41] E. Garfield, Can citation indexing be automated, in: Statistical association methods for mechanized documentation, symposium proceedings, National Bureau of Standards, Miscellaneous Publication 269, Washington, DC, 1965, pp. 189–192.
[42] M.J. Moravcsik, P. Murugesan, Some results on the function and quality of citations, Soc. Stud. Sci. 5 (1975) 86–92.
[43] S. Teufel, J. Carletta, M. Moens, An annotation scheme for discourse-level argumentation in research articles, in: A.L. Henry, S. Thompson (Eds.), Proceedings of the Ninth Conference On European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Bergen, Norway, 1999, pp. 110–117.
[44] H. Nanba, M. Okumura, Towards multi-paper summarization using reference information, in: T. Dean (Ed.), Proceedings of Sixteenth International Joint Conference On Artificial Intelligence. IJCAI 99, 31 July-6 Aug. 1999, Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999, pp. 926–931.
[45] S. Teufel, A. Siddharthan, D. Tidhar, Automatic classification of citation function, in: Proceedings of the 2006 conference on empirical methods in natural language processing, Association for Computational Linguistics, 2006, pp. 103–110.
[46] M. Hernandez-Alvarez, J.M. Gomez S, Citation impact categorization: for scientific literature, in: J.C. Ferreira (Ed.), IEEE 18th International Conference On Computational Science and Engineering (CSE), 21-23 Oct. 2015, IEEE Computer Society, Los Alamitos, CA, USA, 2015, pp. 307–313.
[47] H. Nanba, M. Okumura, Towards multi-paper summarization using reference information, in: IJCAI, 1999, pp. 926–931.
[48] T. Nakagawa, K. Inui, S. Kurohashi, Dependency tree-based sentiment classification using CRFs with hidden variables, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2010, pp. 786–794.
[49] C. Jochim, H. Schutze, Towards a generic and flexible citation classifier based on a faceted classification scheme, in: Proceedings of the 24th International Conference on Computational Linguistics, COLING (COLING 2012 Organizing Committee, Mumbai, India, 2012, pp. 1343–1358.
[50] A. Meyers, Contrasting and corroborating citations in journal articles, in: RANLP, 2013, pp. 460–466.
[51] C.-T. Tsai, G. Kundu, D. Roth, Concept-based analysis of scientific literature, in: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, ACM, 2013, pp. 1733–1738.

[52] M. Abdullatif, Y.S. Koh, G. Dobbie, S. Alam, Verb selection using semantic role labeling for citation classification, in: Proceedings of the 2013 workshop on Computational scientometrics: theory & applications, ACM, 2013, pp. 25–30.
[53] A. Di Iorio, A.G. Nuzzolese, S. Peroni, Characterising citations in scholarly documents: the CiTalO framework, in: Extended Semantic Web Conference, Springer, 2013, pp. 66–77.
[54] M. Abdullatif, Y.S. Koh, G. Dobbie, Unsupervised semantic and syntactic based classification of scientific citations, in: International Conference on Big Data Analytics and Knowledge Discovery, Springer, 2015, pp. 28–39.
[55] J. Donahue, L.A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 677–691.
[56] S. Li, W. Li, C. Cook, C. Zhu, Y. Gao, A Fully Trainable Network with RNN-based Pooling, CoRR, abs/1706.05157 (2017).
[57] M. Liang, X. Hu, Recurrent convolutional neural network for object recognition, 2015), pp. 3367–3375.
[58] S. Li, W. Li, C. Cook, C. Zhu, Y. Gao, Independently Recurrent Neural Network (IndRNN): Building A Longer and Deeper RNN, CoRR, abs/1803.04831 (2018).
[59] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[60] R. JHA, A.-A. JBARA, V. QAZVINIAN, D.R. RADEV, NLP-driven citation analysis for scientometrics, Nat. Lang. Eng. (2017) 1–38.
[61] G. Balikas, S. Moura, M.-R. Amini, Multitask learning for fine-grained twitter sentiment analysis, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, Shinjuku, Tokyo, Japan, 2017, pp. 1005–1008.
[62] H. Fu, Z. Niu, C. Zhang, H. Yu, J. Ma, J. Chen, Y. Chen, J. Liu, ASELM: Adaptive semi-supervised ELM with application in question subjectivity identification, Neurocomputing 207 (2016) 599–609.
[63] J.K. Tarus, Z. Niu, A. Yousif, A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining, Future Generation Comp. Syst. 72 (2017) 37–48.
[64] S. Wan, Z. Niu, An e-learning recommendation approach based on the self-organization of learning resource, Knowl.-Based Syst. 160 (2018) 71–87.

Abdallah Yousif received his BSc with Computer Science and information technology in 2007 from Al-Neelain University and his MSc degree in Computer Sciences from University of Gezira, Sudan in 2010. He is currently a PhD student at School of Computer Science and Technology, Beijing Institute of Technology, China. He also works in Wad Medani Technical Collage, Sudan Technological University as a lecturer, his research interests include Sentiment analysis, machine learning, citation analysis,and data mining.

Zhendong Niu received his Ph.D degree in Computer Science from Beijing Institute of Technology in 1995. He is a professor and deputy dean of the School of Computer Science and Technology, Beijing Institute of Technology, China. He is an adjunct professor in the School of Computing and Information at University of Pittsburgh from 2006. His research areas include digital libraries, e-learning techniques, information retrieval and recommender systems. He serves as an editorial board member for international journal of learning technology. He has published more than 180 papers in journals and international conferences in his field.



James Chambua received his BSc with Computer Science in 2004 and MSc in Computer Science in 2009 from the University of Dar-es-Salaam. He is currently pursuing Ph.D. in Computer Science at the Beijing Institute of Technology (BIT). He works as an Assistant Lecturer at the Computer Science and Engineering Department of the University of Dar-es-Salaam. His research interests include Machine Learning, data mining algorithms, recommender systems and e-learning techniques.

Zahid Younas Khan received his MSc degree and MPhil degree of Computer Science from Quaid-i-Azam University Islamabad in 2005 and 2015, repectively. He is currently pursuing Ph.D. in Computer Science at Beijing Institute of Technology (BIT). He works as a Lecturer at the department of Computer Science and Information Technology, University of Azad Jammu & Kashmir. His research interests include Machine Learning, Data Mining algorithms and recommender systems.