MF-Cite : Citation Intent Classification in Scientific Papers Based on Multi-Feature Fusion

Anonymous ACL submission

Abstract

Citations are crucial in scientific works. Citation analysis techniques help in literature search, citation recommendation, scientific assessment and other research works. Citation intent classification has proved to be useful as an important branch of citation analysis techniques, which categorizes the role that citations play in research works. However, scientific papers usually contain words that are difficult to understand and semantically uncertain, while we find that the classification labels have a greater relationship with the part-of-speech properties of the words in the citation context. Therefore, in this work, we propose a scientific text classification model called MF-Cite that combines citation context feature, WordNet[1]-based semantic feature, and part-of-speech feature. It fuses them for scientific text representation, enabling the model to enhance the understanding of specialized domain terms and accurately comprehend the grammatical information of sentences. Experiments show that our method achieves more favorable results on the ACL-ARC and SciCite datasets.

1 Introduction

In recent decades, the explosive growth of scientific papers has made citation analysis techniques very important. Generally, authors cite literature with the purpose of borrowing existing research backbround to support their work (Gilbert, 1977). Citations play different roles in the citation context and involve different citation intents, such as describing background information about a study or comparing the results of a paper with those of other work (Varanasi et al., 2021). Citation intent classification helps to measure the impact of papers, venues, researchers, etc., or to understand the development and evolution of a field (Jurgens et al., 2018). Additionally, it can provide basic data for literature retrieval and recommendation to build more accurate literature retrieval systems (Berrebbi et al., 2022). In this paper, we utilize WordNet external knowledge base (Fellbaum and Miller, 1998) to supplement the semantic information of the words in the citation context. Then we use SciBERT (Beltagy et al., 2019) pre-trained model and GCN (Scarselli et al., 2009) to extract the contextual and structural information of the citation context and its semantic feature, and finally fuse them. Figure 1 shows the correlation between the classification labels and some words in the citation context, usually nouns, verbs and adjectives have a greater impact on the classification results. Therefore, we add part-of-speech feature to allow the model to capture the syntactic relations and semantic roles in the sentence, which helps to accurately understand the grammatical structure and context of the sentence.

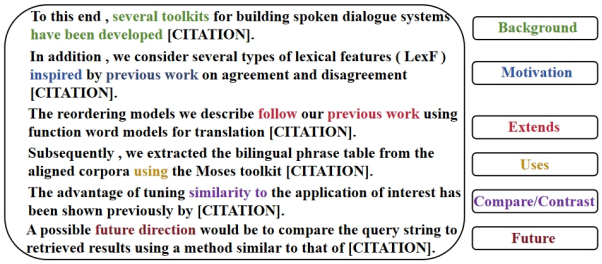


Figure 1: The effect of word parts-of-speech on labels in the context citation. The labels used for illustration here are from the annotation schema proposed for the ACL-ARC dataset (Jurgens et al., 2018).

The contributions of this paper are summarized as follows:

- We obtain the representation information of citation context by fusing contextual information, WordNet-based semantic information and part-of-speech information.
- We propose a new model, MF-Cite, which utilizes SciBERT and GCN to extract infor-

[1] https://wordnet.princeton.edu/

mation from contextual and semantic features and interaction module to deliver and filter the information between them.

- We also compare the model to various baselines, achieving SOTA results on the ACL-ARC dataset and comparable results on the SciCite dataset (Cohan et al., 2019).

The rest of the paper is organized as follows. Sections 2 and 3 of the paper provide related works and some preliminary content about the concepts used in the work. Section 4 describes the design of the model. Section 5 details the experimental setup and results. The paper concludes with Section 6, giving conclusions and other tasks to which the model can be applied.

2 Related Work

In this section, we briefly introduce the existing works of citation intent classification systems and automatic classification methods.

2.1 Citation Intent Classification Systems

In the last few years, there are a number of classification systems that are widely recognized and used. (Jurgens et al., 2018) proposed citation intent classes, using a large number of features such as pattern-based features, topic-based features and prototypical argument features, and a random forest classifier for prediction. (Cohan et al., 2019) proposed a much larger dataset, SciCite, that classified citation intent into three coarse-grained labels. The 3C citation context classification task (Kunnath et al., 2020) was organized by The Open University, UK. They used a portion of the multidisciplinary dataset ACT (Pride and Knoth, 2020) as a dataset, it can be used as a benchmark for future research. (Kunnath et al., 2022) had since extended the 3C dataset further to enrich the features of citing and cited papers.

2.2 Automatic Classification Methods

Automatic methods for citation intent classification have been evolving. (Varanasi et al., 2021) used SciBERT-based multi-task learning to classify citation intent, with the main task being citation intent classification and the three auxiliary tasks being citation worthiness, section title and cited paper title. (Zhang et al., 2022) used native information related to citation context, such as section name and paper title to improve the performance of the model. (Berrebbi et al., 2022) constructed citation graphs that include papers, authors and venues, and co-prediction with citation context significantly improved state-of-the-art results. (Budi and Yaniasih, 2022) used a convolutional neural network to represent citation contexts in journal articles from Indonesia in five scientific fields. (Lahiri et al., 2023) proposed a method based on prompt learning to transform the classification task into a completive prediction task. (Gupta et al., 2023) combined citation contexts and their neighboring contextual sentences and proposed a Transformer-based deep neural network that fuses peripheral sentences and domain knowledge.(Shi et al., 2024) proposed a prompt learning method based on data augmentation and L2 regularization to classify scientific text.

Most of the previous works relied on additional data or information related to citation context to provide more useful information for citation intent classification and obtain promising results. Unlike them, in order to better represent scientific text and classify citation intent, this paper starts from the citation context itself, and utilizes the WordNet external knowledge base to obtain its synonymous sentence as semantic feature, as well as to obtain the part-of-speech feature of the citation context, and finally fuses them. GCN is also utilized to effectively fuse the syntactic information of the sentences to obtain richer textual representation.

3 Preliminaries

In this section, we introduce some preliminary works such as the definition of citation intent classification task and dependency parsing.

3.1 Problem Definition

Formally, a citation intent classification training dataset can be denoted as D = {X,Y}, where X is the instance set and Y is the citation intent label set. Each instance $x \in$ X consists of several tokens $x = [w_1, w_1, \ldots, w_n]$ along with a class label $y \in$ Y. An example of Y is the set {"Background", "Uses", "Compare/Contrast", "Motivation", "Extends", "Future"}. We present the task as estimating the conditional probability $Pr(y|x)$ based on the training set, and identifying the class label to which citation context belongs by $y^{'} = argmax_{y \in Y} Pr(y|x)$.

3.2 Dependency Parsing

Dependency parsing, also known as dependency syntactic parsing, serves to identify interdepen-

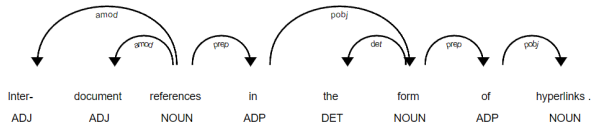


Figure 2: An example of dependency tree.

dencies between words in a sentence. (Robinson, 1970) proposed that the other constituents of a sentence are subordinate to a particular constituent, assuming that the master-slave relationship between words is binary and unequal. In this paper, we use the Spacy toolkit [2] for dependency parsing of sentences. Figure 2 shows the results of dependency parsing for the citation context of the sentence "Inter-document references in the form of hyperlinks." in the ACL-ARC dataset. As we can see, the arrow points from the center word to the dependent word, which must depend on the center word for its existence.

4 Method

In this section, we propose a novel model named MF-Cite for citation intent classification. The overall framework of the proposed model is depicted in Figure 3. It uses SciBERT and GCN for feature extraction of citation context feature and its semantic feature respectively. It utilizes the crosstext-decoder to extract the relationship between the two feature vectors. And a gated network to fuse the features obtained by the crosstext-decoder while filtering out the redundant information. MF-Cite uses the Glove pre-trained model (Pennington et al., 2014) to embed the part-of-speech feature of the citation context, and finally integrates them.

4.1 Obtain Features

Contextual feature: Citation context feature in ACL-ARC and SciCite datasets, it mainly refer to texts containing citation symbols that cite relevant scientific literature.

Semantic feature: The diversity of word meanings in scientific papers makes it more difficult for the model to understand the semantic information of citation context. we use WordNet external knowledge base to obtain the synonymous sentence of citation context, and extend the semantic information of citation context by integrating WordNet knowledge. WordNet organizes nouns, verbs, adjectives and adverbs into a synonym network, so in citation context we obtain the synonyms with the highest semantically similar to these words as semantic supplementary units.

Part-of-speech feature: In citation context, the words that have more influence on classification labels are usually nouns, verbs and adjectives, so integrating part-of-speech information into text representation helps to understand citation context and extract information from citation context. This study uses the "pos_tag" module of NLTK [3] to obtain part-of-speech feature.

4.2 Encoder Layer

For contextual and semantic features, we use SciBERT pre-trained model and GCN to extract information. SciBERT is a variant of BERT pre-trained model (Devlin et al., 2019). The corpus for SciBERT training is scientific papers in the biomedical as well as computer science direction, which makes it easier to understand the semantics expressed in scientific papers (Beltagy et al., 2019). It stacks l layers of h-head self-attention mechanism. Define Given a sentence $x = [w_1, w_1, \ldots, w_n]$ as a sentence of n words, we use SciBERT model to encode each word w_i into a word vector, denoted as $h_i \in \mathbb{R}^d$,where d denotes the dimension of the word embedding generated by SciBERT. Then we denote the vector of sentence s as $\mathbf{H} \in \mathbb{R}^{l \times n \times d}$. Specifically, $\mathbf{H}_l \in \mathbb{R}^{n \times d}$ presents the hidden representation matrix of the last layer in $\mathbf{H}$. The multi-head self-attention matrix of each layer stores the global interrelationship as the distribution of attention between tokens, which can be denoted as $\mathbf{A} \in \mathbb{R}^{l \times h \times n \times n}$, $\mathbf{A}_l \in \mathbb{R}^{h \times n \times n}$ denotes the last layer's multi-head self-attention matrix. Thus, for citation context feature, the hidden representation matrix of the last layer is denoted as $\mathbf{H}_l^c$, and the multi-head self-attention matrix of the last layer is denoted as $\mathbf{A}_l^c$. For semantic feature, the hidden representation matrix of the last layer is denoted as $\mathbf{H}_l^s$, and the multi-head self-attention matrix of the last layer is denoted as $\mathbf{A}_l^s$.

According to the topology of the graph, GCN obtains the embedding vectors of the augmented nodes by aggregating the neighbor information of the nodes. The graph can be represented as G = (V, E), where V denotes the words that make up the text, and E denotes the edge between words. In this paper, the hidden representation of the last layer of each token in SciBERT is regarded as a node v

[2] https://spacy.io/

[3] https://www.nltk.org/

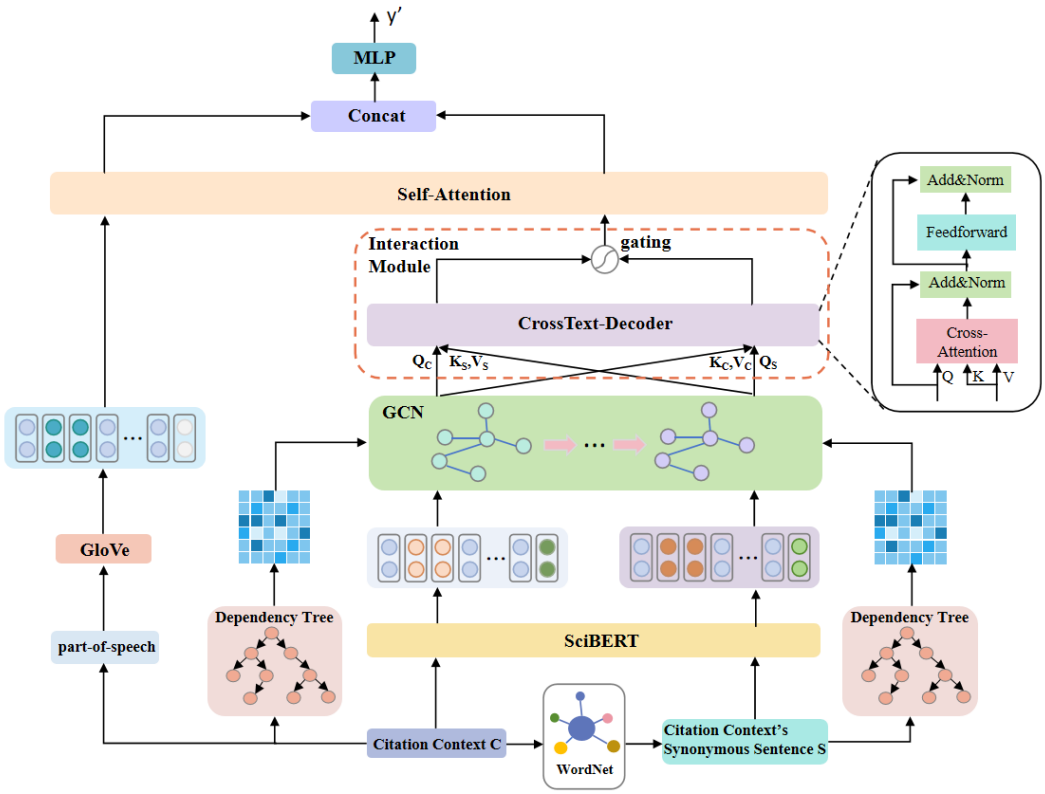


Figure 3: The overview architecture of our proposed MF-Cite model.

in the graph, and the edge e is computed from the dependency adjacency matrix and the multi-head self-attention matrix.

We perform dependency parsing on contextual and semantic features, and transform the obtained dependency tree $\mathbf{T}$ into a dependency adjacency matrix $\mathbf{D}$.

$$\mathbf{D}_{ij} = 1 \qquad if\ \mathbf{T}(w_i, w_j) \tag{1}$$

where $\mathbf{D}_{ij}$ is the dependency of words w_i and w_j in the sentence. To retain more information, we construct an undirected graph $\mathbf{D}_{ij} = \mathbf{D}_{ji} = 1$, and set a self-loop $\mathbf{D}_{ii} = 1$. We obtain the dependency adjacency matrix $\mathbf{D}^c \in \mathbb{R}^{n\times n}$ for citation context and the dependency adjacency matrix $\mathbf{D}^s \in \mathbb{R}^{n\times n}$ for semantic feature.

We compute the mean of all heads of the last layer of the multi-head self-attention matrix for contextual and semantic features, respectively.

$$\begin{aligned} \bar{\mathbf{A}}_l^c &= \frac{1}{h}\sum_{i=1}^{h}\mathbf{A}_l^c \\ \bar{\mathbf{A}}_l^s &= \frac{1}{h}\sum_{i=1}^{h}\mathbf{A}_l^s \end{aligned} \tag{2}$$

where $\bar{\mathbf{A}}_l^c \in \mathbb{R}^{n\times n}$ is the mean value of the attention of all heads in the last layer about the contextual feature, and $\bar{\mathbf{A}}_l^s \in \mathbb{R}^{n\times n}$ is the mean value of the attention of all heads in the last layer about the semantic feature.

To incorporate dependency parsing information, we define the adjacency matrix $\mathbf{M}$ of the graph as the Hadamard product of the self-attention matrix $\mathbf{A}$ and the dependency adjacency matrix $\mathbf{D}$:

$$\begin{aligned} \mathbf{M}^c &= \bar{\mathbf{A}}_l^c \odot \mathbf{D}^c \\ \mathbf{M}^s &= \bar{\mathbf{A}}_l^s \odot \mathbf{D}^s \end{aligned} \tag{3}$$

where $\mathbf{M}^c$ is the adjacency matrix of contextual feature and $\mathbf{M}^s$ is the adjacency matrix of semantic feature.

Next, the GCN uses dependency paths to transform and propagate information between paths and aggregates the propagated information to update the node embeddings. The contextual node initial embedding $\mathbf{N}_0^c = \mathbf{H}_l^c$ and semantic node initial embedding $\mathbf{N}_0^s = \mathbf{H}_l^s$.

$$\begin{aligned} \mathbf{N}_{k+1}^c &= \tau(\widetilde{Z}_c^{-\frac{1}{2}}\mathbf{M}^c\widetilde{Z}_c^{-\frac{1}{2}}\mathbf{N}_k^c W_k^c) \\ \mathbf{N}_{k+1}^s &= \tau(\widetilde{Z}_s^{-\frac{1}{2}}\mathbf{M}^s\widetilde{Z}_s^{-\frac{1}{2}}\mathbf{N}_k^s W_k^s) \end{aligned} \tag{4}$$

where $\mathbf{N}_{k+1}^c$ is the embedding of the k+1th layer of the contextual node and $\mathbf{N}_{k+1}^s$ is the embedding of the k+1th layer of the semantic node. $\widetilde{Z}$ is the

degree matrix of the adjacency matrix $\mathbf{M}$,defined as $\widetilde{Z} = \widetilde{Z}_{ii} = \sum_j \mathbf{M}_{ij}$. W_k is the weight matrix of the kth layer. τ is the ReLU activation function.

After the last node features update, we obtain the contextual node representation $\mathbf{N}_c$ and the semantic node representation $\mathbf{N}_s$, which contain syntactic information. For part-of-speech feature, we use the Glove pre-trained model for vector representation, denoted as $\mathbf{N}_p$.

4.3 Interaction Layer

In order to extract the relationship between contextual and semantic features, to transfer and interact information between citation contexts and their synonymous sentences, and help the model better understand the semantic information of words, we use a part of the decoder in the Transformer model proposed by (Vaswani et al., 2017) for the computation.

First, the contextual feature vector $\mathbf{N}_c$ is computed as the query vector and the semantic feature vector $\mathbf{N}_s$ is computed as the key vector and the value vector. Then the semantic feature vector $\mathbf{N}_s$ is denoted as the query vector and the contextual feature vector $\mathbf{N}_c$ is denoted as the key vector and value vector. The multi-head attention values between them are calculated separately. The following is the method to compute a head in the multi-head attention.

$$\begin{aligned} head^i_{(N_c,N_s)} &= softmax(\frac{Q^i_c {K^i_s}^\top}{\sqrt{d_k}})V^i_s \\ head^i_{(N_s,N_c)} &= softmax(\frac{Q^i_s {K^i_c}^\top}{\sqrt{d_k}})V^i_c \end{aligned} \tag{5}$$

where $Q^i_c = \mathbf{N}_c W^i_{Q_c}$, $K^i_s = \mathbf{N}_s W^i_{K_s}$, $V^i_s = \mathbf{N}_s W^i_{V_s}$. $W^i_{Q_c}$ is the weight matrix of Q_c when calculating the ith head. $W^i_{K_s}$ is the weight matrix of K_s when calculating the ith head. $W^i_{V_s}$ is the weight matrix of V_s when calculating the ith head. Q^i_s, K^i_c, V^i_c are calculated as above.

As Equation (6) shows, each head is concatenated to calculate the multi-head attention:

$$\begin{aligned} MultiHead_{(N_c,N_s)} &= Concat(head^1_{(N_c,N_s)}, \ldots, \\ &\quad head^n_{(N_c,N_s)})W_{(N_c,N_s)} \\ MultiHead_{(N_s,N_c)} &= Concat(head^1_{(N_s,N_c)}, \ldots, \\ &\quad head^n_{(N_s,N_c)})W_{(N_s,N_c)} \end{aligned} \tag{6}$$

where $W_{(N_c,N_s)}$ and $W_{(N_s,N_c)}$ are weight matrixs. n denotes the number of attention heads.

Next, the remaining steps of the crosstext-decoder are computed using contextual feature as an example. Semantic feature is computed in a similar way. After calculating the cross-attention between contextual feature and semantic feature, the new contextual feature vector representation $\mathbf{N}^{'}_c$ is obtained by residual connection and layer normalization.

$$\mathbf{N}^{'}_c = LayerNorm(\mathbf{N}_c + MultiHead_{(N_c,N_s)}) \tag{7}$$

Subsequently, $\mathbf{N}^{'}_c$ is fed into a second sublayer that includes feedforward neural network, residual connection and layer normalization operation. Afterwards the decoder output $\widehat{\mathbf{N}}_c$ of the contextual feature can be obtained.

$$\begin{aligned} \mathbf{N}^{''}_c &= FeedForward(\mathbf{N}^{'}_c) \\ \widehat{\mathbf{N}}_c &= LayerNorm(\mathbf{N}^{'}_c + \mathbf{N}^{''}_{\mathbf{c}}) \end{aligned} \tag{8}$$

The decoder output $\widehat{\mathbf{N}}_s$ of semantic feature is computed as above.

Finally, we design the gated network to filter out redundant information while fusing contextual feature $\widehat{\mathbf{N}}_c$ and semantic feature $\widehat{\mathbf{N}}_s$.

$$\begin{aligned} \mathbf{N}_{gate} &= \sigma((W_1\widehat{\mathbf{N}}_c + W_2\widehat{\mathbf{N}}_s) \odot \widehat{\mathbf{N}}_c) \\ &\quad +(1 - \sigma((W_1\widehat{\mathbf{N}}_c + W_2\widehat{\mathbf{N}}_s) \odot \widehat{\mathbf{N}}_s)) \end{aligned} \tag{9}$$

where $\mathbf{N}_{gate}$ is the output of the gated network. σ denotes the sigmoid activation function. W_1 and W_2 are weight matrixs. $\odot$ denotes Hadamard product.

The interaction module is designed to encourage the information to flow between different features. With cross-attention, each token can combine information from another feature, enabling the extension of word semantics in the citation context. In addition, the gated network filters out redundant information to ensure that the final representations are relevant to the current task.

4.4 Self-attention Layer

In order to better learn the information of each feature itself, we use the self-attention mechanism to calculate the importance score of the current token and the remaining tokens, and then get the final weighted sum representations.

$$\begin{aligned} Att_{gate} &= softmax(\frac{Q_{gate}{K_{gate}}^\top}{\sqrt{d_k}})V_{gate} \\ Att_p &= softmax(\frac{Q_p {K_p}^\top}{\sqrt{d_k}})V_p \end{aligned} \tag{10}$$

Categories	Count
Background	1021
Uses	365
Compare/Contrast	344
Motivation	98
Extends	73
Future	68

Table 1: Details of ACL-ARC dataset.

Categories	Count
Background	6376
Method	3153
Result comparison	1491

Table 2: Details of Scicite dataset.

where Att_{gate} denotes the self-attention vectors of $\mathbf{N}_{gate}$, Att_p denotes the self-attention vectors of the part-of-speech feature $\mathbf{N}_p$.

4.5 Prediction Layer

Finally, the attention vectors of the fusion features of contexts and semantics and the attention vectors of the part-of-speech feature are concatenated to predict the probability of each label of the citation context.

$$y^{'} = \sigma(mlp([Att_{gate}; Att_p])) \quad (11)$$

where σ denotes the sigmoid activation function.

5 Experiments

This section evaluates the performance of our proposed MF-Cite. We conduct extensive experiments on two widely-used citation intent classification datasets and provide a comprehensive comparison with existing baselines.

5.1 Datatsets

We use two standard citation intent classification datasets, **ACL-ARC** and **SciCite**. **ACL-ARC** (Jurgens et al., 2018) consists of 186 papers from the ACL Anthology Reference Corpus (Bird et al., 2008) and contains 1,941 instances labeled with 6 citation intent labels. Table 1 provides detailed information about ACL-ARC's label sets and instances. **SciCite** (Cohan et al., 2019) is a much larger dataset built from 6,627 papers and has 11,020 instances tagged with 3 categories of coarse-grained citation intents. Table 2 provides information about SciCite's label sets and instances.

5.2 Baseline Models

We compare our model with previous strong baselines for citation intent classification. **RF** (Jurgens et al., 2018) denotes the random forest classifier based on a variety of manually-selected features — structural, lexical and grammatical, field and usage features — that signify different aspects of a scientific paper. **Structural-Scaffold** (Cohan et al., 2019) is a multi-task learning framework containing BiLSTM with attention, using GloVe word embedding vectors concatenated with ELMo (Peters et al., 2018) as the input. **SciBERT Fine-tune** (Beltagy et al., 2019) is fine-tuned SciBERT pre-trained language model and uses SciBERT as the input of a fully connected layer to make predictions. **MPMAF** (Qi et al., 2022) is defined as a citation intent classification method based on MPNet pre-training and multi-head attention feature fusion, where inputs are citation context feature and citation external features consisting of grammatical word frequency feature and citation structure feature. **GraphCite** (Berrebbi et al., 2022) takes citation context feature and citation graph features containing papers, authors and venues together to make predictions. **MTCIC** (Qi et al., 2023) considers the correlation between citation intents, citation sections and citation worthiness classification tasks, and build a multi-task citation classification framework with soft parameter sharing constraint. **CitePrompt** (Lahiri et al., 2023) proposes the prompt learning method for citation intent classification by selecting the appropriate pre-trained language model, the prompt template and the prompt verbalizer. **Aug-L2 Prompt** (Shi et al., 2024) uses prompt tuning based on data augmentation and L2 regularization to classify scientific text.

5.3 Implementation Details

In the experiments, we first remove all special characters and lower cases for the citation contexts. We freeze a portion of the SciBERT weight parameters for 768-dimensional embedding, and the GCN model also has embeddings of size 768. The cross-text decoder is a 2-layer stacked structure. And the loss function is computed by using cross-entropy. We utilize the AdamW optimizer (Loshchilov and Hutter, 2017), whose L2 regularization parameter is 0.01. The learning rate of the two datasets is 2e-5 and batch_size is 8. The epoch is 10 for the ACL-ARC dataset and 5 for the SciCite dataset.

Method	F1	P	R
RF	54.60	64.90	49.90
Structural-Scaffold	67.90	81.30	62.50
SciBERT Finetune	71.70	73.28	72.12
MPMAF	72.80	79.82	70.08
GraphCite	77.34	78.65	78.79
MTCIC	75.78	82.07	72.80
MF-Cite(Ours)	**81.04**	**84.59**	**78.96**

Table 3: Macro results (F1, Precision, Recall) on the ACL-ARC dataset.

Method	F1	P	R
RF	79.20	82.80	77.80
Structural-Scaffold	84.00	84.70	83.60
SciBERT Finetune	86.15	**87.86**	84.95
CitePrompt	86.33	-	-
Aug-L2 Prompt	85.88	-	-
MTCIC	85.79	86.48	85.20
MF-Cite(Ours)	**87.66**	87.17	**88.20**

Table 4: Macro results (F1, Precision, Recall) on the SciCite dataset.

All experiments are implemented with Pytorch[4] and trained on NVIDIA RTX 3090 24GB GPU.

5.4 Results and Analysis

5.4.1 Main Results

Table 3 shows the main experimental results of the model proposed in this paper on the ACL-ARC dataset with several baselines. Firstly, we observe that the MF-Cite model proposed in this paper achieves noticeable improvements over the state-of-the-art method on the citation intent classification task. Compared to RF model, our model increases 26.44%, 19.69%, and 29.06% on each metric, respectively. Also we find that the pre-trained model can significantly improve the scores on each metric. For GraphCite, the previous state-of-the-art model on macro-F1 and recall, our model improves by 3.7% and 0.17%, respectively. Our model outperforms MTCIC, the previous state-of-the-art model on precision score, by 2.52%.

Table 4 shows the main experimental results of the model proposed in this paper on the SciCite dataset with several baselines. Our model achieves state-of-the-art results on most of the evaluation metrics. The macro-F1 score is 1.33% higher than the model with the previous state-of-the-art results, and the recall score is 3% higher than the model with the previous state-of-the-art results. However, it did not perform as well as the fine-tuned SciBERT on precision score. Meanwhile, we find that our method performs better on macro-F1 than prompt learning.

5.4.2 Ablation Study

To further validate the effect of different input features and modules on model performance, we conduct some ablation experiments on the ACL-ARC and SciCite datasets.

Input feature. Table 5 shows the effect of adding semantic and part-of-speech (pos) features successively on the model. Adding semantic feature on top of citation context, both datasets increase on all three evaluation metrics, where the ACL-ARC dataset improves by 9.66%, 5.28%, and 9.5%, respectively. Meanwhile, we find that the addition of part-of-speech information has a positive impact on the results of the model. Compared to the SciCite dataset, the ACL-ARC dataset has a larger floating improvement, we suppose that it may be due to the smaller size of the ACL-ARC dataset. For the small-scale dataset, with the increase of the number of features, the amount of information learned by the model and the complexity of the model have a greater impact on the results. Semantic feature has a greater impact on the ACL-ARC dataset, while for the SciCite dataset part-of-speech feature plays a bit more of a role.

Module effect. Table 6 shows the effect of removing the interaction module and the self-attention module on the model, respectively. We find that the interaction module is more important for both datasets. The ACL-ARC dataset has a higher decrease in model performance after removing the interaction module. The possible reason is that the model doesn't filter out redundant information, too much noise is more sensitive for small size dataset, which results in poorer model performance. The SciCite dataset shows a decrease in performance by removing the self-attention module by 0.71%, 1.22%, and 0.13%.

5.4.3 Visualization

The confusion matrix shown Figure 4 shows the nature of the errors made by our model. In ACL-ARC, the model makes more errors in recognizing the "Motivation" and "Extends" labels, mislabeling these as "Background", probably because most of the times the purpose of the citation is to provide relevant information about the domain. The imbal-

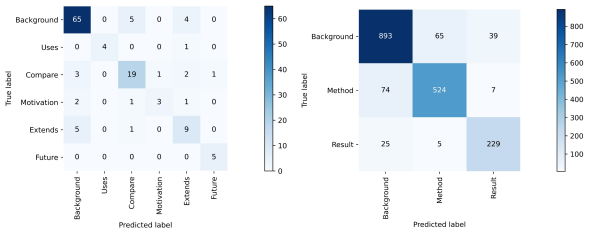
[4] https://pytorch.org/

Method	ACL-ARC			SciCite		
	F1	P	R	F1	P	R
context	67.72	74.00	66.36	85.52	84.46	87.53
context w / semantic	77.38	79.28	75.86	86.33	85.69	87.56
context w / semantic,pos	**81.04**	**84.59**	**78.96**	**87.66**	**87.17**	**88.20**

Table 5: Influence of different input features on the model.

Method	ACL-ARC			SciCite		
	F1	P	R	F1	P	R
w/o interaction	68.65	74.09	64.72	85.11	83.99	86.97
w/o self-attention	74.44	78.46	72.15	86.95	85.95	88.07
MF-Cite(Ours)	**81.04**	**84.59**	**78.96**	**87.66**	**87.17**	**88.20**

Table 6: Influence of different modules on the model.



(a) ACL-ARC (b) SciCite

Figure 4: Confusion matrix results on two datasets.

ance of the dataset makes it easy for the model to recognize other labels as "Background". In the SciCite dataset, 12.23% of the true labels "Method" and 9.65% of the true labels "Result" are misclassified as "Background".

6 Conclusion

In order to realize automatic classification of citation intent, in this paper, we propose a scientific text classification model called MF-Cite. It integrates citation context feature, WordNet-based semantic feature and part-of-speech feature. This model first fuses contextual feature with semantic feature to extend the semantic information of words in context and enhance the model's understanding of specialized domain terms. Because part-of-speech contains important semantic information and different part-of-speech have different importance for label, we integrate the part-of-speech information with them to improve the scientific text representation ability of the model.

Experimental results suggest that the MF-Cite model proposed in this paper outperforms the contrast methods. We further validate the effectiveness of individual input feature and module. Finally our model can also be applied to scientific text representation tasks such as academic paper rating (Xue et al., 2023), citation recommendation (Lu et al., 2023), and citation count prediction (Li et al., 2019; van Dongen et al., 2020).

Ethical Statement

In this paper, we use available data in all experiments. No relevant data is released publicly, and the model trained on this dataset doesn't present any new or greater risks. They are not dangerous to humans. Therefore, we do not find any ethical issues.

Limitations and Future Work

We summarize the limitations of our method as follows: 1)To some extent, part-of-speech information of citation context has an effect on the classification labels. However, we simply add part-of-speech information in our model, and this choice is not optimal. In the future, we can increase the part-of-speech weight vector and assign greater weight to the part-of-speech (nouns, verbs, adjectives) that contributes more to the classification labels. We plan to integrate the part-of-speech feature of each word and its weight feature into the text representation of scientific papers to improve the model effect. 2)This paper uses the datasets that contain only one sentence of citation context and one label, whereas the citation context in paper may involves multi-

ple sentences and the citation plays more than one role (Lauscher et al., 2022). Therefore, in future work, we plan to use a citation context dataset that contains multiple sentences and multiple labels.

References

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 3613–3618. Association for Computational Linguistics.

Dan Berrebbi, Nicolas Huynh, and Oana Balalau. 2022. GraphCite: Citation intent classification in scientific publications via graph embeddings. In *Companion Proceedings of the Web Conference 2022*, pages 779–783.

Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. 2008. The ACL Anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA).

Indra Budi and Yaniasih Yaniasih. 2022. Understanding the meanings of citations using sentiment, role, and citation function classifications. *Scientometrics*, 128:735–759.

Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. 2019. Structural scaffolds for citation intent classification in scientific publications. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3586–3596. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Christiane Fellbaum and George Miller. 1998. *WordNet : an electronic lexical database*. MIT Press.

G. Nigel Gilbert. 1977. Referencing as persuasion. *Social Studies of Science*, 7(1):113–122.

Priyanshi Gupta, Yash Kumar Atri, Apurva Nagvenkar, Sourish Dasgupta, and Tanmoy Chakraborty. 2023. Inline citation classification using peripheral context and time-evolving augmentation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 3–14.

David Jurgens, Srijan Kumar, Raine Hoover, Daniel A. McFarland, and Dan Jurafsky. 2018. Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, 6:391–406.

Suchetha Nambanoor Kunnath, David Pride, Bikash Gyawali, and Petr Knoth. 2020. Overview of the 2020 WOSP 3C citation context classification task. In *Proceedings of the 8th International Workshop on Mining Scientific Publications*, pages 75–83. Association for Computational Linguistic.

Suchetha Nambanoor Kunnath, Valentin Stauber, Ronin Wu, David Pride, Viktor Botev, and Petr Knoth. 2022. ACT2: A multi-disciplinary semi-structured dataset for importance and purpose classification of citations. In *International Conference on Language Resources and Evaluation*, pages 3398–3406. European Language Resources Association.

Avishek Lahiri, Debarshi Kumar Sanyal, and Imon Mukherjee. 2023. Citeprompt: Using prompts to identify citation intent in scientific papers. *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 51–55.

Anne Lauscher, B. R. Ko, Bailey Kuehl, Sophie Johnson, David Jurgens, Arman Cohan, and Kyle Lo. 2022. MultiCite: Modeling realistic citations requires moving beyond the single-sentence single-label setting. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1875–1889. Association for Computational Linguistics.

Siqing Li, Wayne Xin Zhao, Eddy Jing Yin, and Ji rong Wen. 2019. A neural citation count prediction model based on peer review text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4914–4924. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Yonghe Lu, Meilu Yuan, Jiaxin Liu, and Minghong Chen. 2023. Research on semantic representation and citation recommendation of scientific papers with multiple semantics fusion. *Scientometrics*, 128:1367–1393.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237. Association for Computational Linguistics.

David Pride and Petr Knoth. 2020. An authoritative approach to citation classification. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, pages 337–340.

Ruihua Qi, Zhen Shao, Jinghua Guan, and Xu Guo. 2022. Citation intent classification method based on mpnet pretraining and multi-head attention feature fusion. *Pattern Recognition and Artificial Intelligence*, 35(9):849–857.

Ruihua Qi, Jia Wei, Zhen Shao, Zhengguang Li, Heng Chen, Yunhao Sun, and Shaohua Li. 2023. Multi-task learning model for citation intent classification in scientific publications. *Scientometrics*, 128:6335 – 6355.

Jane J. Robinson. 1970. Dependency structures and transformational rules. *Language*, 46(2):259–285.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80.

Shijun Shi, Kai Hu, Jie Xie, Ya Guo, and Huayi Wu. 2024. Robust scientific text classification using prompt tuning based on data augmentation with l2 regularization. *Information Processing & Management*, 61(1):103531.

Thomas van Dongen, Gideon Maillette de Buy Wenniger, and Lambertus Schomaker. 2020. SChuBERT: Scholarly document chunks with BERT-encoding boost citation count prediction. In *Proceedings of the First Workshop on Scholarly Document Processing*, pages 148–157. Association for Computational Linguistics.

Kamal Kaushik Varanasi, Tirthankar Ghosal, Piyush Tiwari, and Muskaan Singh. 2021. IITP-CUNI@3C: Supervised approaches for citation classification (task a) and citation significance detection (task B). In *Proceedings of the Second Workshop on Scholarly Document Processing*, pages 140–145. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 5998–6008.

Zhi Xue, Guoxiu He, Jiawei Liu, Zhuoren Jiang, Star X. Zhao, and Wei Lu. 2023. Re-examining lexical and semantic attention: Dual-view graph convolutions enhanced bert for academic paper rating. *Information Processing & Management*, 60(2):103216.

Yang Zhang, Rongying Zhao, Yufei Wang, Haihua Chen, Adnan Mahmood, Munazza Zaib, Wei Emma Zhang, and Quan Z. Sheng. 2022. Towards employing native information in citation function classification. *Scientometrics*, 127:6557–6577.