Citation intent classification using POS features

Tejaswi Choppa
M.Sc. Computational Linguistics
st180670@stud.uni-stuttgart.de

Sheethal Velutharambath
M.Sc. Computational Linguistics
st176435@stud.uni-stuttgart.de

Abstract

Citations are used for informative purposes, such as providing background information, demonstrating the application of methodology, or comparing results with other publications. In this work, the aim is to identify the intent of a citation, whether it is background, method, or result. We used SciCite data which includes 11K manually annotated citation intentions in the computer science and biomedical disciplines based on citation context. Here we are using six models, support-vector machine (SVM), perceptron, SGDClassifier, Long short-term memory (LSTM) and Bidirectional Long short-term memory (BiLSTM)) with part-of-speech (POS) features. We evaluated the models using precision, recall, F1 score, and accuracy metrics. Our main classification architecture focuses on encoding POS tags with Glove Embeddings and BiLSTM for the intent classification. Also, a comparison is made across models to see if POS tags help in improving the performance of the models.

1 Introduction

Citations are an essential component of scientific literature, serving as a means of acknowledging the varied contributions of diverse scientific works Nicholson et al. (2021). Citations are an important part of scientific publishing because they connect research findings throughout time. A citation informs your readers that particular content in your paper came from another scientific paper and provides them with the information they need to locate that scientific paper again. Citations are highly useful for anyone interested in learning more about your ideas and where they came from and certain purposes, such as providing background information, demonstrating the application of methodology, or comparing results with other publications.The task of determining the intent of a given citation is known as citation intent classification.(Small, 2018) Figure 1 is an example of a citation context with an article cited highlighted. The cited paper, which in this case clearly identifies the citation intent of the referenced publication.

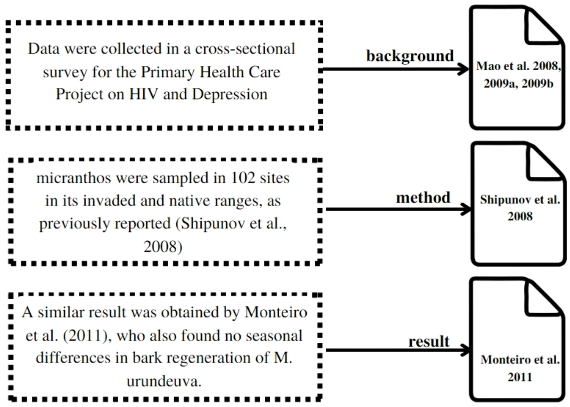

Figure 1: Example of citations with different intents (*Background, Methods,* and *Result*)

(Cohan et al., 2019) illustrate how structural features may be utilized to efficiently guide citation intent categorization. They employ a multitask learning framework with two scaffolds related to the work of predicting citation intent. On the ACL-ARC dataset, their model obtains a state-of-the-art result (F1 score of 67.9 percent) with a 13.3 absolute gain over the best previous results. They introduce SciCite, a new huge dataset of citation intentions, and demonstrate the efficacy of their method.

(Roman et al., 2021) discovered that contextual embedding might play an important role in categorizing the citation sentences, since contextual sentences outperformed non-contextual sentences. They annotated the C2D un-annotated dataset and constructed a new dataset on top of it after evaluating their proposed technique. They solely utilized citation intent classes from the SciCite dataset, such as Background, Method, and Result. (Nicholson et al., 2021) citation indexes aid in measuring the links between scientific works, but they fall short

Intent categories	Definition	Example
Background	Background is used to describe citations that provide further context for a topic, concept, or other component of a field of study or research.	4 lead to a decrease in SC absorption in mice Its utilization in the context of cluster ensembles for gene expression data 18 is another significant future research.
Method	Using a method, tool, strategy, or dataset	We used an active contour algorithm [10] to segment the organs from 340 coronal slices over the two patients.
Result	Comparison of the outcomes of the study with the outcomes of other papers	We compare SISI with the dynamic programming algorithm, temporarily called k effector, in 21 when the infection process follows the IC model.

Table 1: The description and examples of SciCite dataset citation intent types. Adapted from Cohan et al. (2019)

since they do not transmit contextual information about a citation. To address this issue, they created scite, a "smart citation index" that categorizes citations depending on context. Scite was created by analyzing over 25 million full-text scientific papers and now includes a database of over 880 million categorised citation assertions. (Jurgens et al., 2018) citation functions contribute to the reception of an article. A detailed examination of citation demonstrates that writers cite works for a variety of reasons, including background, motivation, extension, usage, contrast, and future. These styles of citation assist us in understanding the current state of research endeavors and their progress across entire scientific domains such as NLP. They link all of this in this study by constructing a cutting-edge classifier for exposing scientific framing and a new corpus annotated with citation function.(Pham and Hoffmann, 2003)

Here we used the SciCite dataset (Cohan et al., 2019) a collection containing 11K carefully annotated citation intentions in the computer science and biomedical disciplines, depending on citation context. The goal is to accurately categorize citation intentions into one of three categories:*Background, Methods,* and *Result*.

The key research topic addressed by this paper was whether part-of-speech can improve the performance of the classifier. The technique of assigning a word in a text to a corresponding part-of-speech tag based on its context and meaning is known as POS tagging. In the word2vec model for creating an embedding sequence, we used part-of-speech and then trained with an SVM classifier. We utilize the scikit-learn package, which simplify our work. We tested our model using Precision, Recall, F1 score and accuracy.

2 Methodology

In the following we introduce the proposed methodology for the task of citation intent classification in research papers. We describe different approaches, namely the traditional machine learning classifiers+ Word2Vec embeddings and Bi-LSTM GloVe embeddings model with and without POS tags, respectively.

2.1 GloVe embeddings

GloVe generates word vector representations using unsupervised learning techniques. It is trained on global word-to-word statistics from big text corpora, and the generated vectors exhibit extremely intriguing linear vector space relations. GloVe is thought to combine the advantages of the Word2Vec skip-gram model in the word analogy. GloVe's benefit is that it does not rely on local data or contextual knowledge about words, but instead integrates worldwide statistics. It comes in a variety of sizes; we utilized the (glove.6B.300d). The tokens have been encoded in the citation context of size n as X = $\{x_1, x_2, \ldots, x_n\}$, where x_i is a word vector that concatenates the non-contextual word representations (Pennington et al., 2014)

2.2 Word2Vec Embeddings

Word2vec is an open source tool and is based on deep learning. It is used to learn the vector representations of words. Word2vec consists of a two-layer neural net that processes the texts. One is the CBOW (continuous bag-of-words) model and another is the continuous skip-gram model that are the two main learning models used in Word2vec. The CBOW model utilizes context to predict a target word and the skip-gram model uses a word to predict a target context. After passing the training text corpus as the input, Word2vec creates vectors that are distributed numerical representations of word features, features such as the context of individual words. Each word is represented by the vector and in this case the Word2Vec has been used to represent the POS tags collected over the citations from the train dataset.

2.3 POS Tags

Part of speech (POS) tagging is the process of assigning the part of speech tag corresponding to the words in a sentence. The words in the description of the citation are assigned POS tags using the nltk POS tagger. These POS tags are helpful as many of the words belonging to the result class correspond to the VB class and likewise. spaCy, a free open-source library for Natural Language Processing that facilitates NER, POS tagging, dependency parsing and word vectors generation has been for POS tagging. The features have been generated in the format of token_tag. (Peters et al., 2018) The table 2 gives an example of the features generated.

2.4 Bidirectional LSTM (BiLSTM)

It simply concatenates forward and backward passes of regular LSTMs. BiLSTMs aid in the utilization of both left and right context information, which is normally required for proper analysis. In the unidirectional LSTM model, only information from the preceding words is captured, as tokens are fed into the network in a feed forward manner. However, in the BiLSTM model, by the input is processed in both forward and backward directions enabling to keep track of information from both directions. (Reimers and Gurevych, 2017)

For a given input sentence X = $\{x_1, x_2, \ldots, x_t\}$, the hidden states for the feed-forward and backward LSTM outputs are calculated as $\vec{h_t}$ and $\overleftarrow{h_t}$, respectively and concatenated to be used as a word representation of each word token in step t.

Token	**Intent**
Study_NN	Background
Mitochondrial_JJ	Method
Comparing_VBZ	Result

Table 2: Features generated after adding POS tags

3 Experiments

In this section we cover the details of training the POS list Embeddings and their Evaluation.

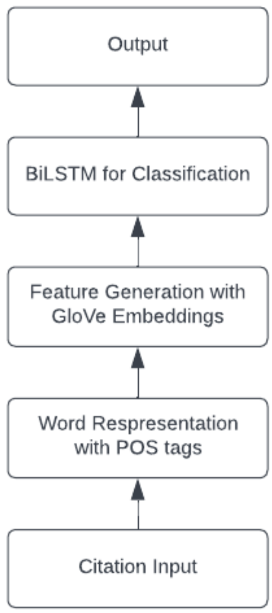

Figure 2: Proposed model for citation intent classification

3.1 Dataset

For the data, the Scicite dataset (Cohan et al., 2019) has been used that consists of 11K carefully annotated citation intentions in the computer science and biomedical disciplines, depending on citation context. In train there are 8,194 instances, in test 1,859 instances and validation 916 instances. Each of these citations corresponds to one of the intent from background, method used and the results compared.Table 1 shows three intent categories: *background, method*, and *result*. Each citation data has

Model	with POS tags	without POS tags
SVM	0.481	0.507
KNN	0.517	0.483
Perceptron	0.508	0.507
SGD Classifier	0.574	0.533
LSTM	0.59	0.563
BiLSTM + GloVe Embeddings	0.77	0.783

Table 3: Accuracy of different models with and without POS tags

been pre-processed, punctuation and stop words removed, stemmed and lemmatized and tokenized using the nltk and spaCy libraries.

3.2 Experimental Setup

The extracted tokens have been assigned a POS tag using the spaCy POS tagger for the LSTM and BiLSTM models and nltk_pos tagger for the other machine learning based classifier models . The extracted POS tags have then been encoded into their corresponding numerical values using the Word2Vec model as described in the section 3. These vectors have then been represented as sentence vectors by sequencing and padding. These vectors essentially form the training data set on which each of our machine learning models are trained. For the classification task, six models, namely the SGD Classifier, SVM, KNN, Perceptron, BiLSTM have been used and compared their performance respectively. For the main architecture, we chose the BiLSTM model for POS features. The experiments for the BiLSTM model, vectors were generated from glove embeddings on POS tagged data in the form of token_tag. The embedding vectors used in the model are initialized to a maximum sequence length of 300. Adam optimizer has been utilized to optimize parameters with a learning rate of 0.001. To prevent the model from over-fitting, early stopping has been employed to monitor the validation loss and stop training if no improvement is achieved for 20 epochs.(Srivastava et al., 2014) The model has been evaluated using the performance metrics accuracy, precision, recall and F1score.

4 Results

In the proposed approach, we noted the effect of using POS tags for all the words in the dataset. We tested our experiments with with different classification models. Different embeddings were employed in combination with the classifiation models and their performance has been observed. Table 3 shows the comparison of the accuracy given by different models with and without the POS tags. There are inconsistencies in the performance of the models where the accuracy shows very little improvement by including the POS tags. Also the F1 score for the BiLSTM + GloVe embeddings is 0.765 and 0.76 with and without POS tags, which shows that many of the POS tags within different classes overlap and couldn't help in classifying distinguishably.

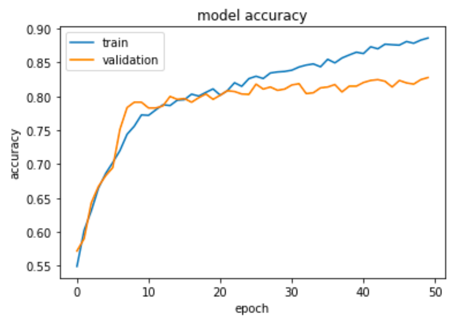

Figure 3: Training model Accuracy for the proposed POS BiLSTM model

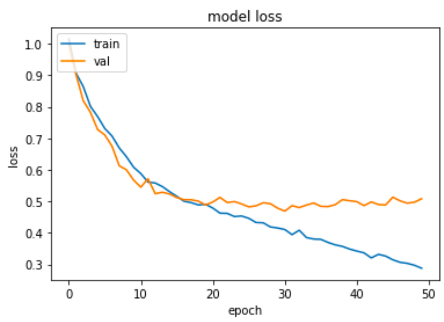

Figure 4: Training model Loss for the proposed POS BiLSTM model

5 Conclusion and Future Work

Although a little improvement is observed in POS tagged models compared to the non POS tagged models, the improvement is not that significant. However, this could be mainly due to employing POS tags for the entire vocabulary which makes it slightly difficult for the tags to be significant determiners for the classes. However, we would like to propose that in future, the model can be trained using POS tags focusing on few key words which play a crucial role in predicting the class. Also, the employment of n-grams with the POS tags can seemingly increase the performance of the model.

6 Contributions

Abstract, Introduction, Dataset by Sheethal Methodology, Experiments, Results, Conclusion by Tejaswi[1].

References

[1] 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, Minneapolis, Minnesota. Association for Computational Linguistics.

[2] David Jurgens, Srijan Kumar, Raine Hoover, Dan 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.

[3] Josh M. Nicholson, Milo Mordaunt, Patrice Lopez, Ashish Uppala, Domenic Rosati, Neves P. Rodrigues, Peter Grabitz, and Sean C. Rife. 2021. scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*, 2(3):882–898.

[4] Jeffrey Pennington, Richard Socher, and Christopher 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, Doha, Qatar. Association for Computational Linguistics.

[5] 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, New Orleans, Louisiana. Association for Computational Linguistics.

[1] https://github.com/SheethalVelutharambath/citation_intent

[6] Son Bao Pham and Achim Hoffmann. 2003. A new approach for scientific citation classification using cue phrases. pages 759–771.

[7] Nils Reimers and Iryna Gurevych. 2017. Reporting Score Distributions Makes a Difference: Performance Study of LSTM-networks for Sequence Tagging. *arXiv e-prints*, page arXiv:1707.09861.

[8] Muhammad Roman, Abdul Shahid, Shafiullah Khan, Anis Koubaa, and Lisu Yu. 2021. Citation intent classification using word embedding. *Ieee Access*, 9:9982–9995.

[9] Henry Small. 2018. Characterizing highly cited method and non-method papers using citation contexts: The role of uncertainty. *Journal of Informetrics*, 12(2):461–480.

[10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958.