

# On the Role of Syntactic Graph Convolutions for Identifying and Classifying Argument Components

Gaku Morio, Katsuhide Fujita

Tokyo University of Agriculture and Technology  
 2-24-16, Nakacho, Koganei, Tokyo, Japan. +81423887003.  
 morio@katfujiiab.tuat.ac.jp, katfujii@cc.tuat.ac.jp

## Abstract

This paper focuses on fundamental research that combines syntactic knowledge with neural studies, which utilize syntactic information in argument component identification and classification (AC-I/C) tasks in argument mining (AM). The following are our paper’s contributions: 1) We propose a way of incorporating a syntactic GCN into multi-task learning models for AC-I/C tasks. 2) We demonstrate the valid effectiveness of our proposed syntactic GCN in fair experiments in some datasets. We also found that syntactic GCNs are promising for lexically independent scenarios. Our code in the experiments is available for reproducibility.<sup>1</sup>

## Introduction

Argument mining (AM) is an emerging discipline that focuses on extracting claims and premises and infers their structures from discourses. There are two fundamental tasks in AM: argument component identification and argument component classification (AC-I/C) (Stab and Gurevych 2017). AC-I focuses on the separation of argumentative and non-argumentative text units and the identification of argument component boundaries. AC-C classifies argument components into different types, such as claims and premises.

Recent several years have seen the technical advances of neural AM. Most recent studies on AC-I/C tasks employ word knowledge (i.e., word embeddings and recurrent neural networks), while syntactic knowledge also should play an important role for AC-I/C tasks. This insight is derived from the fact that syntactic knowledge is robust for context independent argument component classifications (Lippi and Torroni 2015). However, to the best of our knowledge, few studies have been proposed that utilize syntactic information in AC-I/C tasks for neural techniques.

Hence, this paper focuses on fundamental research that combines syntactic knowledge with neural studies. We propose a syntactic graph convolutional network (GCN) for AC-I/C tasks. We conducted fair experiments with our model and existing models to show the genuine effectiveness of GCNs. After combining LSTMs and CRFs, our re-

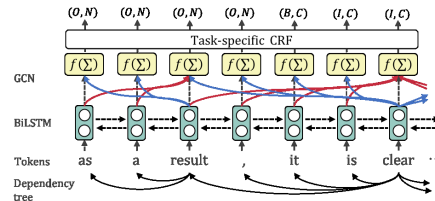


Figure 1: Our syntactic GCN with MTL. Red arrows show the opposite convolution, and blue arrows show the along convolution.

sults show that GCN improved the AC-I/C tasks for some datasets.

## Problem Settings

Our study focuses on two key problem settings. First, AC-I/C tasks are considered sequence tagging problems. We predict token-level BIO tags (AC-I) and argument component types (AC-C). In other words, each input token has a gold label:  $\{(b, t) \mid b \in \{B, I, O\}, t \in Types\}$ , where  $b$  denotes whether the current token is non-argumentative ( $O$ ) or whether it begins ( $B$ ) or continues ( $I$ ) a component;  $t$  is the component type, i.e., claim ( $C$ ), premise ( $P$ ), or non-argumentative ( $N$ ).

Second, we provide multi-task learning (MTL), which improves the performance of the AM problem. This paper uses a state-of-the-art MTL (Schulz et al. 2018), which employs bi-directional LSTM (BiLSTM) and conditional random field (CRF) layers, namely, a BLC. All of the BiLSTM parameters are shared by all the tasks, and an independent CRF layer is provided for each subtask (AC-I and AC-C).

## Syntactic Graph Convolutional Networks

We propose a BLC-based model with a syntactic GCN (Figure 1). First, the BiLSTM encodes each token by concatenating both of the forward and backward states of the LSTMs. A context-aware representation of a token is produced:  $h_i = \vec{h}_i \circ \overleftarrow{h}_i$ . Second, the encoded token representation produced in the BiLSTM is fed into our syntactic GCN, which is based on dependency convolution (Marcheggiani and Titov 2017). Given each hidden state  $h_i$  from BiLSTM, GCN calculates a

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup><https://github.com/EdoFrank/GCNArgMining>

	Identification (AC-I)			Classification (AC-C)		
	Essay	News	Wiki	Essay	News	Wiki
BLC	83.8/88.3	58.2/61.1	47.4/64.8	67.0/75.5	59.0/60.3	39.3/63.5
BLC_GCN (ours)	<b>84.9/88.9</b>	<b>58.5/61.4</b>	45.1/64.8	<b>67.6/76.0</b>	<b>59.6/61.0</b>	37.4/ <b>64.2</b>

Table 1: Performances in macro- and micro-F1: **Bold** denotes significant differences ( $p < 0.05$ ) using Mann-Whitney U-Test.

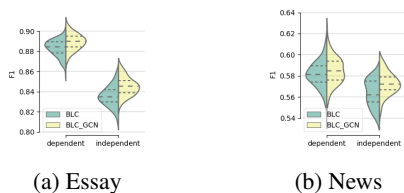


Figure 2: BLC vs. BLC\_GCN in main task: GCN achieves a better performance in lexically independent scenarios.

node (a token) representation and encodes information about its immediate syntactic neighbors:

$$GCN(h_i) = f \left( \sum_{j \in N(i)} g_{i,j} (V_{dir(j,i)} h_j + b_{L(j,i)}) \right),$$

where  $V_{dir(j,i)}$  denotes a direction-specific weight matrix and  $b_{L(j,i)}$  is a label-specific bias. For example, the label of  $L(result, as)$  in Figure 1 is *case*.  $f$  is an activation function (e.g., ReLU).  $N(i)$  are neighbors of token  $i$ .  $g_{i,j}$  is a gating scalar for assigning a weight to each edge.

## Experiments

**Data** We conducted our experiments with three datasets: essay (Stab and Gurevych 2017), news (Habernal et al. 2018), and wiki (Biran and Rambow 2011) for AC-I/C tasks. Although each dataset has a different scale and a different argument component type, token-level segmentation and typing are available.

**Implementation** Each dataset is divided randomly as  $train : test = 8 : 2$ , and then for each train dataset,  $train : develop = 7 : 3$ . These experiments were conducted more than 120 times with a random seed for each model, and the average scores are shown. We selected the model with a maximum development score (a macro F1 of the main task in our case). In addition, 50-dim GloVe embeddings and part-of-speech embeddings are the input representation.

**Results** Table 1 summarizes the results of our model and the BLC baseline. The GCN layers generally improved the scores in the essay and news datasets, indicating that a syntactic GCN has valid effectiveness for AC-I/C tasks in specific datasets. A notable point of this result is that the GCN model yields better performances not only for the main task (AC-I) but also for the auxiliary task (AC-C). However, the wiki dataset’s performance was disappointing since the large

label imbalance of BIO in it, where nearly 70% of the data are ‘O’.

Next, we investigated the mechanism of the syntactic GCN for lexically or partially semantically independent scenarios. For this, experiments with lexically independent input were conducted. GloVe embeddings from the input representation were removed, and only experiments with part-of-speech embeddings were conducted. Figure 2 shows histograms of the essay and news datasets in the micro-F1 of the AC-I task. The lexically independent scenario is hard to learn. However, GCN in turn significantly outperforms the baseline in the lexically independent scenario compared to the lexically dependent scenario. This result implies that syntactic GCNs are effective for lexically independent situations.

## Conclusion and Future Work

This paper proposed the application of a dependency-based syntactic GCN to MTL for AC-I/C tasks. The experiment results indicated that our proposed syntactic GCN outperformed the baselines in predicting argument component boundaries and types for not highly imbalanced datasets. Our experimental result with a lexically independent scenario implied syntactic GCNs are effective with insufficient context information. One possible future work will study GCN’s performance in other situations.

## Acknowledgments

This work was supported by JST AIP-PRISM (JP-MJCR18ZL) and JST CREST (JPMJCR15E1), Japan.

## References

- Biran, O., and Rambow, O. 2011. Identifying justifications in written dialogs. In *Proceedings of ICSC’11*, 162–168.
- Habernal, I.; Wachsmuth, H.; Gurevych, I.; and Stein, B. 2018. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In *Proceedings of NAACL2018, Volume 1*, 1930–1940. ACL.
- Lippi, M., and Torroni, P. 2015. Context-independent claim detection for argument mining. In *Proceedings of IJCAI’15*, 185–191. AAAI Press.
- Marcheggiani, D., and Titov, I. 2017. Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of EMNLP2017*, 1506–1515. ACL.
- Schulz, C.; Eger, S.; Daxenberger, J.; Kahse, T.; and Gurevych, I. 2018. Multi-task learning for argumentation mining in low-resource settings. In *Proceedings of the 2018 Conference of the NAACL-HLT, Volume 2*, 35–41. ACL.
- Stab, C., and Gurevych, I. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics* 43(3):619–659.