Constituency Tree Representation for Argument Unit Recognition

Samuel Guilluy, Florian Méhats and Billal Chouli

Université Rennes - IRMAR

Argument and Discourse Theory

- An Elementary Discourse Unit (EDU) is a concept used in the field of discourse analysis to refer to a minimal unit of meaning within a larger discourse or conversation. It represents a self-contained piece of information that contributes to the overall structure and coherence of a discourse.
- An Argument Discourse Units (ADU) are units of meaning that contribute to the development and presentation of an argumentative structure. ADUs typically contain propositions, claims, evidence, or reasoning that support or challenge a particular standpoint or claim. They can be expressed as individual sentences, groups of sentences, or even longer stretches of text.

In practical applications, while certain studies rely on the annotator's judgment to determine the boundary of an ADU, many studies prefer to utilize a set of syntactic rules as a foundation. This approach is favored because employing syntactic structure for annotating a sizable corpus at the token level is comparatively easier.

Evaluation of the node similarity

We have established three metrics to evaluate the suitability of employing the constituency tree representation for argument unit recognition.

- Leaf-Leaf similarity metric: This refers to the ratio of nodes in a linear chain sentence representation that have both the same label and are adjacent to each other. In the cases where a constituency tree representation is available (columns 4, 5, 6), we further narrow down this set of nodes to those that not only share the same label but also belong to the same grammatical class, meaning they have the same parent node.
- Leaf-Interior Node (IN) similarity metric: This indicates the proportion of leaf nodes that share the same label as their corresponding grammatical class.
- IN-IN similarity metric: This measures the ratio of internal nodes that are connected by an edge and have the same label.

Metrics	Dataset	No tree	Depth 2	Depth 3	Depth 4
Leaf-Leaf similarity	AURC	97.1 %	98.3%	98.4%	98.6%
	CDCP	97.8 %	99.6 %	99.7 %	99.7%
	ESSAYS	91.9 %	97.9%	0%	0%
	ARG2020	95.4 %	97.9%	98.2 %	98.6 %
Leaf-IN similarity	AURC	//	90.2%	91.3%	84.2%
	CDCP	//	98.1%	98.7%	68.9%
	ESSAYS	//	89.2 %	0%	0%
	ARG2020	//	91.8%	92.9%	59.2%
IN-IN similarity	AURC	//	88.5%	93.3%	92.9%
	CDCP	//	96.4%	97.8%	90.4%
	ESSAYS	//	85.1 %	0%	0%
	ARG2020	//	91.7%	95.1%	88.3%

Table 1. Evaluation of three proportions to assess the appropriateness of using the constituency tree representation. ("IN" means Interior Node.)

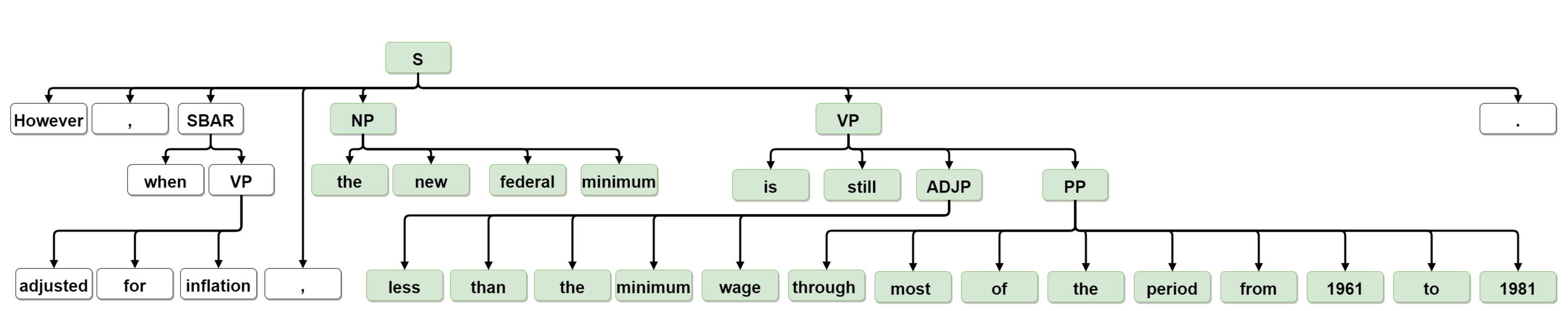

Figure 1. Constituency tree representation of the sentence, according to the Universal POS tags categories (with a depth of the tree cap at three): "However, when adjusted for inflation, the new federal minimum is still less than the minimum wage through most of the period from 1961 to 1981." about the topic "minimum wage". The green nodes represents words or spans with "PRO" label and the grey nodes represents words and spans with "NON" label.

Datasets Presentation

- Argument Unit Recognition and Classification (AURC) [5] is an argument mining corpus annotated with argumentative structure information capturing the polarity of arguments given a topic. It is composed of 8000 sentences, equally divided into 8 topics. Given a topic (such as "School uniforms" or "abortion"), the authors distinguished between PRO (supporting), CON (opposing) arguments and NON (non-argumentative) words. In order to construct the sentence level label of, they use the following rule : if only NON occurs, the sentence is labeled NON. If NON and only PRO (resp. only CON) occurs, PRO (resp. CON) is chosen. If both PRO and CON occur, the more frequent label of the two is assigned.
- The Cornell eRulemaking Corpus (CDCP) [3] is an argument mining corpus annotated with argumentative structure information capturing the evaluability of arguments. The corpus consists of 731 user comments on Consumer Debt Collection Practices (CDCP) rule by the Consumer Financial Protection Bureau (CFPB); the resulting dataset contains 4931 elementary unit and 1221 support relation annotations.
- ARG2020 [1] is an argument mining corpus annotated with argumentative structure composed of "claims" and "premises". It is composed of 145 essays English argumentative essays selected through the Writing Mentor Educational App. It is based on middle school students writing. The claims is defined as a potentially arguable statement that indicates a person is arguing for or arguing against something. The premises is defined as the reasons given by either for supporting or attacking the claims.

Models Presentation

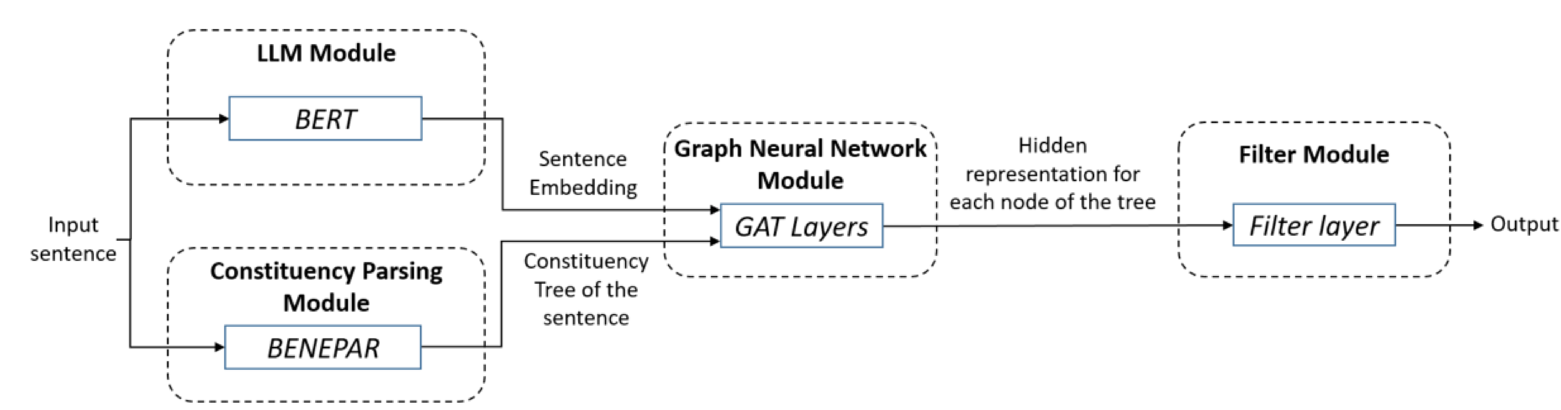

Figure 2. Illustration of the architecture of the proposed model incorporating a filter layer.

Our model architecture, as depicted in Figure 2, consists of four modules.

- LLM module: We use the pretrained BERT model This choice aligns with the original approach taken in the reference article [5], allowing us to compare our results with their baseline.
- Constituency Parsing module: To construct the constituency tree, we employed the Berkeley Neural Parser (BENEPAR) [2]. We assigned labels to the internal nodes (NT) of the constituency tree to capture their representations. Since these annotations were not available in the dataset, we opted to annotate the internal nodes using the same labeling rules described earlier for sentence labeling. This decision ensures the internal logic of the labeling process is preserved.
- Graph Neural Network module: We choose Graph Attention Network (GAT) model in order to learn the relation between nodes using the attention mechanism.
- Filter module: The multi-layer RNN module, known as the filter layer module, is designed to run on each individual tree filter. Its purpose is to facilitate information propagation exclusively among nodes residing within the same filter.

Results

The results corresponding to the token level classification are presented in Table 2. We reached our best results with the model composed of BERT - GNN - Filter Layer when we allowed a maximal depth of 3. The major improvement of our model is when we evaluate at token level. This is due to the better identification of the ADU when we constraint the model to consistency tree.

Model	F1 score at token level		
	AURC	CDCP	ARG2020
BERT	68 %	80 %	55 %
BERT - Linear chain CRF	69 %	81 %	55.5 %
BERT - GAT - Linear chain CRF	72.8 %	83 %	58 %
BERT - GAT - Filter Layer	73.5 %	83.5 %	61 %

Table 2. F1- score of the different models at token level

Interpretability

Consider the problem of attributing the model prediction to its input features. First, we use the Integrated Gradient method [4] which computes a score for every input of the model compared to a baseline value. The score is calculated by integrating the gradient along the input representation from the baseline to our input. When this gradient is higher, our input data is more relevant for prediction. Then, we study the importance of the edges for the model prediction. We iteratively hide one edge from the graph and compute the model to see how the result varies. By doing so with all the edges we can measure the necessity of every edges taken independently. An example of the interpretability on edges can be found in Figure 3. It shows that the words which are labeled with an argument (in color) furnish more information to their related edges. This confirm the utility of adding the constituency tree representation to our model.

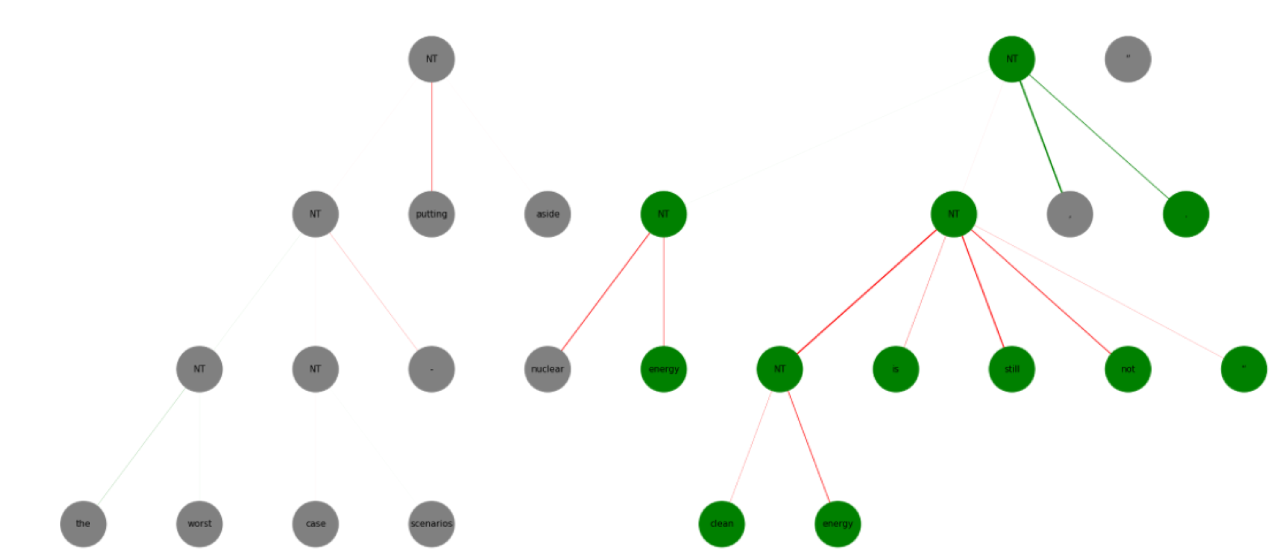

Figure 3. Tree representation of the edges interpretability methods. The red edges refer to edges which have a positive impact on the prediction. The green edges refer to edges which have a negative impact on the prediction.

References

[1] Tariq Alhindi and Debanjan Ghosh.
"Sharks are not the threat humans are": Argument Component Segmentation in School Student Essays.
In Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications, pages 210–222, Online, April 2021. Association for Computational Linguistics.

[2] Nikita Kitaev and Dan Klein.
Constituency Parsing with a Self-Attentive Encoder.
In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2676–2686, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[3] Joonsuk Park and Claire Cardie.
A Corpus of eRulemaking User Comments for Measuring Evaluability of Arguments.
In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).

[4] Mukund Sundararajan, Ankur Taly, and Qiqi Yan.
Axiomatic Attribution for Deep Networks, June 2017.
arXiv:1703.01365 [cs].

[5] Dietrich Trautmann, Johannes Daxenberger, Christian Stab, Hinrich Schütze, and Iryna Gurevych.
Fine-Grained Argument Unit Recognition and Classification.
Proceedings of the AAAI Conference on Artificial Intelligence, 34(05):9048–9056, April 2020.
Number: 05.

samuel.guilluy@univ-rennes.fr