Incorporating Zoning Information into Argument Mining from Biomedical Literature

Boyang Liu, Viktor Schlegel, Riza Batista-Navarro, Sophia Ananiadou

Department of Computer Science, The University of Manchester

INTRODUCTION

Argument mining can be divided into four sub-tasks:

- Argumentative component recognition (separating argumentative units from non-argumentative units)
- Argumentative component classification (classifying argumentative component types, i.e., claim or premise)
- Relation recognition (finding relations between argumentative components)
- Argumentative relation classification (classifying argumentative relation types i.e., support or attack)

Based on granularity, it can be divided into:

- Sentence-level task: the boundary of an argumentative component is the same as a sentence
- Token-level task: the length of argumentative components can range from less than a clause to several sentences.

Background: *We have recently suggested that bolus 5-fluorouracil (5-FU) may work via a RNA directed mechanism while ...*

Patients and methods: *Two hundred fourteen patients from nineteen Italian centers were randomized to the control arm ...*

Results: {*Nine CR and twenty-seven PR were obtained on one hundred eleven evaluable patients treated in experimental arm (RR = 32%, 95% confidence interval (95% CI): 24%-42%), while two CR and eleven PR were observed among one hundred three evaluable patients in control arm (RR = 13%, 95% CI: 7%-21%)* }$_{premise1}$. ... {*Eighty percent of patients receiving second-line chemotherapy in control arm were treated with continuous infusion 5-FU* }$_{premise5}$.

Conclusions: *Alternating,* [*schedule-specific biochemical modulation of FU is more active than ...*]$_{claim1}$. [*However, the overall survival was similar suggesting that alternating bolus and infusional 5-FU upfront may be as effective as giving them in sequence as first- and second-line treatment*]$_{claim2}$.

Figure 1: An abstract from PubMed 11142481. We remove several sentences for brevity. The sequences in curly brackets are premises (pieces of evidence supporting or attacking claims) and in square brackets are claims.

MOTIVATION

Text zoning aims at segmenting a text into zones. Here, each zone differs from others and consists of text parts in terms of a particular function. Biomedical literature abstracts can be divided into five zones: *Background, Objective, Method, Result* and *Conclusion.*

Relevance between zoning and argument:

- Argumentative components mainly exist in the *Result* and *Conclusion* zones,
- premises are more likely to occur in the *Result* zone,
- claims are more likely to occur in the *Conclusion* zone.

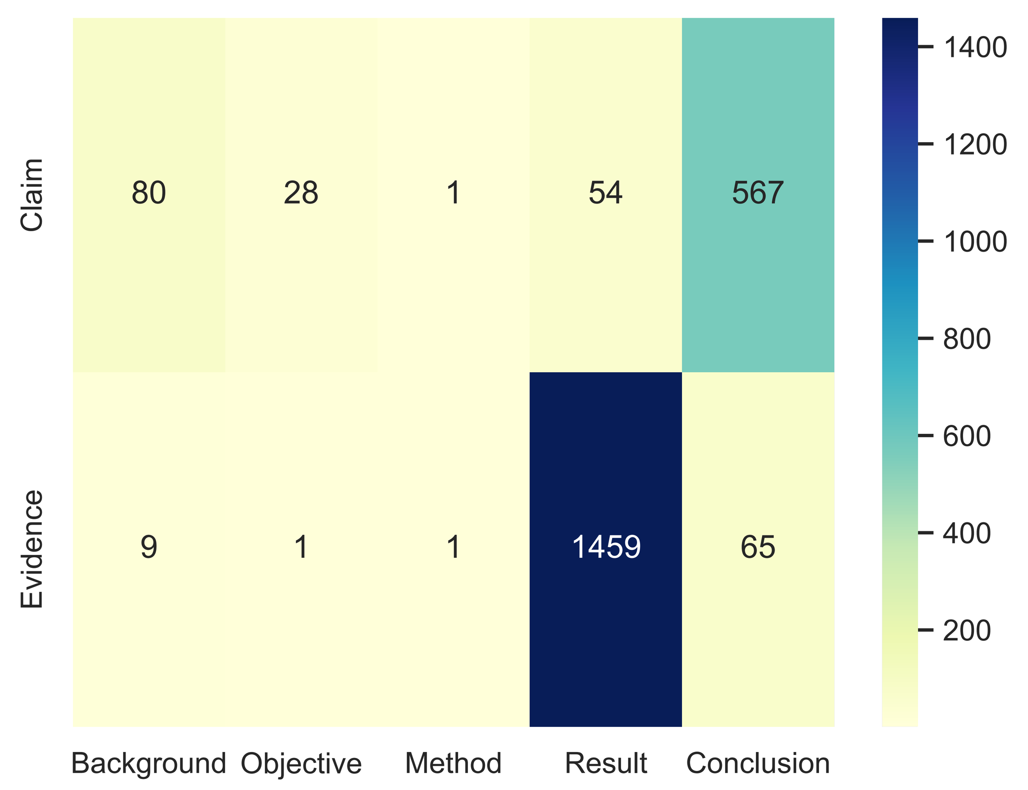

Figure 2: Distribution of argumentative components and zoning information within the training subset of PubMedRCT dataset. Zoning labels are predicted labels using a tool named HSLN (Jin and Szolovits, 2018)

METHODOLOGY

To leverage the zoning information, we placed the corresponding zoning label in front of each sentence in an abstract. The zoning label of each sentence is not golden label but predicted by HSLN (Jin and Szolovits, 2018).

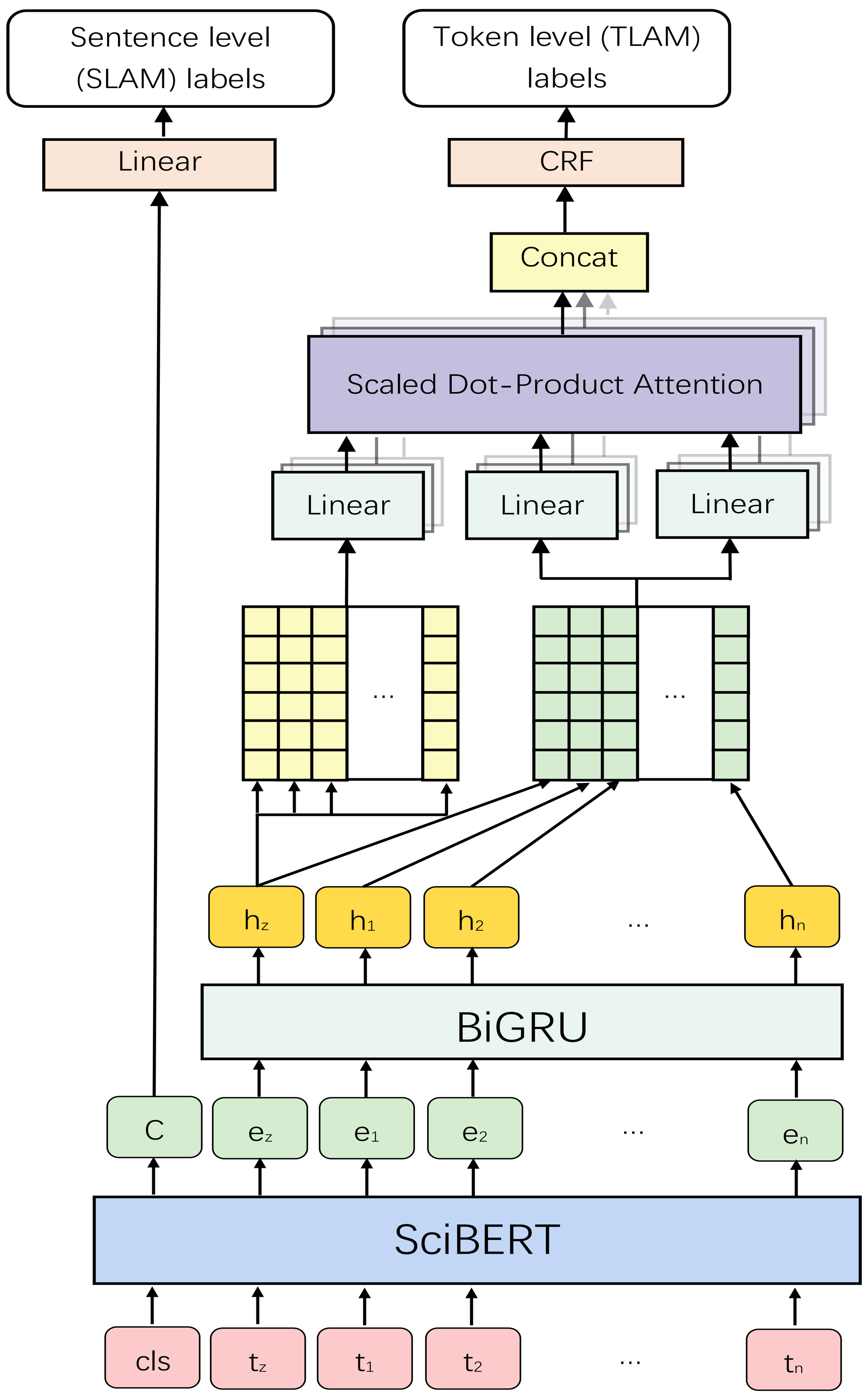

Figure 3: Overview of our model. t_z represents zoning labels, and *cls* is the special token [CLS] in SciBERT

DATASET

AbstRCT dataset (token level):

- Neoplasm corpus, only includes abstracts concerning neoplasm, used for train, development and test.
- Glaucoma corpus only includes abstracts concerning glaucoma, used for test.
- Mixed corpus includes abstracts concerning five diseases (neoplasm, glaucoma, hypertension, hepatitis and diabetes), used for test.

SciARG dataset (sentence-level). It's a fine-grained dataset that contains eleven types of argumentative components (i.e., proposal, observation).

	AbstRCT	SciARG
Number of abstracts	659	285
Number of AC types	3	11
Number of ACs	4198	2787

Table 1: Statistics of the datasets

RESULT

Method	Component	Main Unit
Accuosto et al. (2021) [2]	67.38	86.76
SLAM	**69.08**	**88.79**

Table 2: Results for sentence-level argument mining. Best results are highlighted in bold. SLAM is our sentence-level argument mining model.

RESULT

Heuristic method: Sentences labelled as *Background*, *Objective* and *Method* are classified as non-argumentative sentences, and all the tokens of non-argumentative sentences are all labelled as O (Outside). Sentences labelled as *Result* are considered evidences and labelled as *Conclusion* are classified as claims.

TLAM_Single_{B,O,M,R,C}: The model only exploits a single zoning label, i.e., *Background*, *Objective*, *Method*, *Result* and *Conclusion* respectively.

Models	Neoplasm				Glaucoma				Mixed			
	f1	C-F1	E-F1	F1	f1	C-F1	E-F1	F1	f1	C-F1	E-F1	F1
Heuristic method	87.23	**73.88**	81.96	80.20	87.32	80.04	80.93	82.17	88.00	73.71	83.97	80.97
Mayer et al. (2020) [1]	90.05	69.65	84.17	83.48	91.50	76.52	83.53	85.67	90.88	72.50	83.62	84.27
TLAM	90.78	73.80	86.17	**85.59**	92.18	80.45	85.99	87.82	91.63	75.58	**85.76**	86.05
TLAM_without_Att	90.82	72.94	85.65	85.09	91.93	79.47	84.68	87.03	91.68	74.69	85.33	85.70
TLAM_Single_B	90.12	71.45	85.85	84.64	91.74	80.45	85.62	87.66	90.99	74.63	84.86	85.41
TLAM_Single_O	90.15	72.03	85.79	84.79	91.74	80.38	85.87	87.73	90.94	74.28	84.75	85.25
TLAM_Single_M	90.05	71.89	85.88	84.77	91.79	**80.90**	**86.18**	**87.99**	91.33	74.87	85.37	85.72
TLAM_Single_R	90.68	72.84	86.35	85.29	**92.19**	80.19	85.29	87.52	**91.69**	75.56	85.50	85.98
TLAM_Single_C	**90.94**	73.39	**86.39**	85.51	91.86	80.45	85.45	87.60	91.61	**76.31**	85.38	**86.16**

Table 3: Results for token-level argument mining. All the reported results are statistically significant. Best results are highlighted in bold. TLAM is our token-level argument mining model. F1 and f1 stand for macro- and micro-averaged F1-scores, respectively. C-F1 and E-F1 stand for macro-averaged F1-scores for claim and evidence, respectively.

CONCLUSION

We propose a method leveraging zoning information for the argumentative component identification and classification tasks in the biomedical domain. Specifically, we added the predicted zoning label in front of each sentence, which is then given as input to the encoding layer. Experiment results performed at sentence-level and token-level demonstrate the effectiveness of utilizing zoning information for the task of argument mining.

REFERENCE

1. Accuosto, P., Neves, M., and Saggion, H. (2021). Argumentation mining in scientific literature: from computational linguistics to biomedicine. In Frommholz I, Mayr P, Cabanac G, Verberne S, editors. BIR 2021: 11th International Workshop on Bibliometric-enhanced Information Retrieval; 2021 Apr 1; Lucca, Italy. Aachen: CEUR; 2021. p. 20-36. CEUR Workshop Proceedings.
2. Mayer, T., Cabrio, E., and Villata, S. (2020). Transformer-based argument mining for healthcare applications. In ECAI 2020, pages 2108–2115. IOS Press.
3. Jin, D. and Szolovits, P. (2018). Hierarchical neural networks for sequential sentence classification in medical scientific abstracts. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3100–3109, Brussels, Belgium, October-November. Association for Computational Linguistics.
4. Eger, S., Daxenberger, J., and Gurevych, I. (2017). Neural end-to-end learning for computational argumentation mining. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 11–22, Vancouver, Canada, July. Association for Computational Linguistics.