KU LEUVEN

DMON: A Simple yet Effective Approach for Argument Structure Learning

Speaker:

Wei Sun
PhD candidate in KU Leuven

Outline

- Introduction
  - Motivation
  - Prior work
  - Contribution
- Method
- Ablation study
- Conclusion

Introduction

- Motivation
  - Argument structure learning (ASL) involves detecting and tagging relationships between argumentative components in a text.
    - Example shown below
  - ASL models help facilitate more accurate and deeper comprehension of text and hence play a critical role in various NLP applications
- Prior works
  - Previous works [1, 2, 3] conducted pairwise relation classification for ASL without contextual information.
  - A more recent attempt by [4] encodes the contextual arguments with a transformer architecture. But they still ignored the relationships between contextual arguments.
  - A key challenge posed by argument structure learning, a sub-task of argument mining, is that fully understanding the relationship between two arguments often requires capturing contextual knowledge about other arguments and their relationships.

Introduction

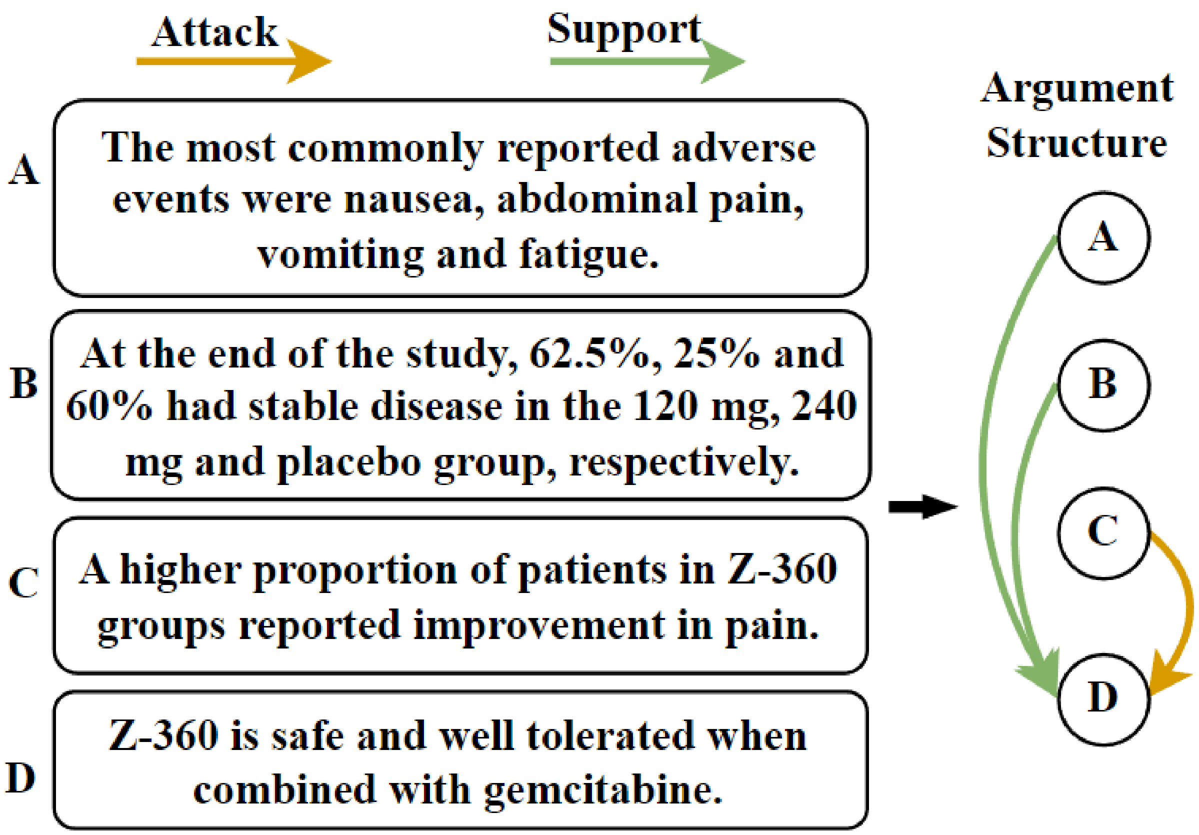


Figure 1: A simple example of an argumentative structure showing attack (orange arrow) and support (green arrow) relationships.

Introduction

- Our Contribution
  - We propose a novel approach called DMON to encode contextual arguments and their relationships by connecting the argument structure with a relationship tensor.
  - We propose a bidirectional learning mechanism that allows distinguishing head and tail arguments in a relationship.
  - We design a cropping strategy to handle the scarcity of training data.
  - Experimental results on five different-domain argument structure learning datasets show that our method outperforms state-of-the-art models for the ASL task.

Method

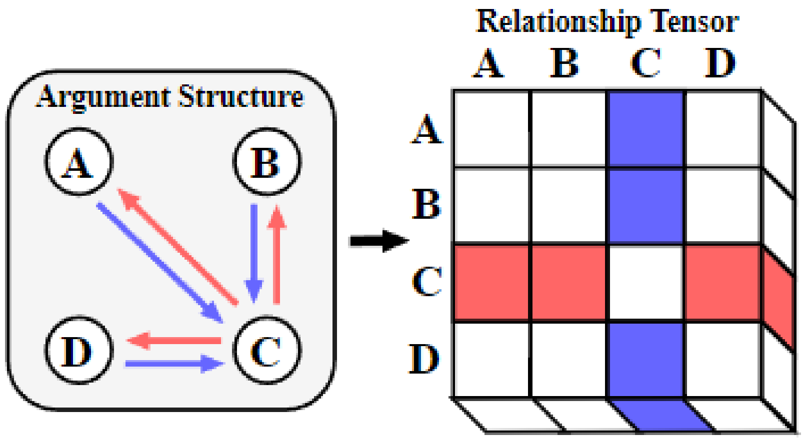


Figure 2: We select argument C as the observation object. This example shows the correlation between the head (red) and tail (blue) relationship information and a relationship tensor. Each element in this relationship tensor is the concatenation of two arguments.

Method

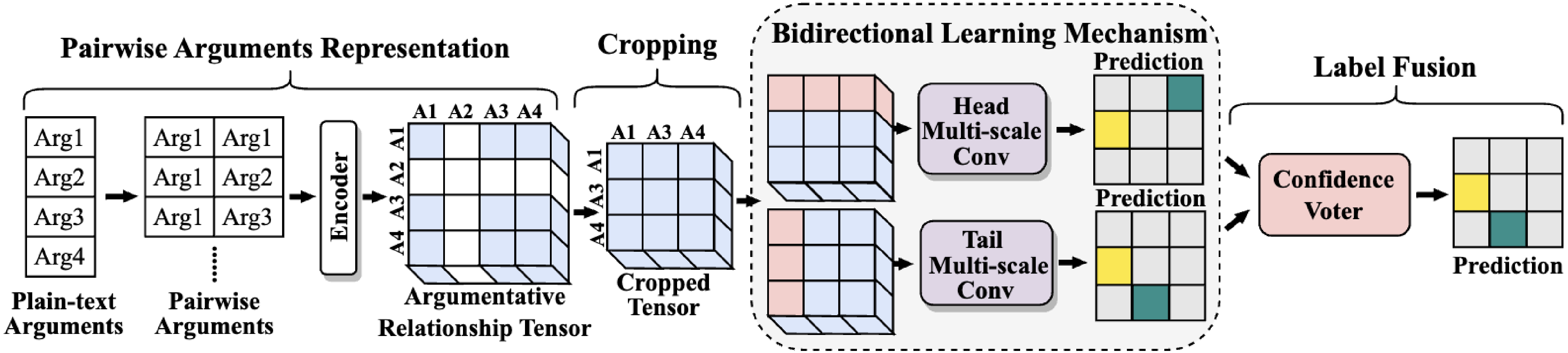


Figure 3. The overall architecture of the DMON model.

Results

Models	Neoplasm				Glaucoma				Mixed			
	F1	S-F1	A-F1	U-F1	F1	S-F1	A-F1	U-F1	F1	S-F1	A-F1	U-F1
AMPERE++	63.73	-	-	-	-	-	-	-	-	-	-	-
Roberta	67.00	-	-	-	66.00	-	-	-	69.00	-	-	-
AMCT-Sci	68.16	59.99	49.12	95.45	62.28	64.71	24.78	95.24	69.43	55.31	58.00	94.76
TransforMED	69.96	58.72	55.65	95.51	69.72	65.32	47.00	96.88	71.82	57.14	**63.41**	94.90
RESATTARG	70.92	52.77	**65.38**	94.54	68.40	54.73	**56.00**	94.36	67.66	49.62	59.09	94.21
DMON	**76.30** ±0.71	**68.25** ±0.98	64.13 ±0.47	**97.33** ±0.08	**74.16** ±1.10	**73.16** ±1.17	53.41 ±0.63	**97.27** ±0.06	**74.07** ±1.11	**68.35** ±1.42	54.05 ±0.74	**97.08** ±0.09

Table 1. Experimental results of argument structure learning in terms of macro-F1 scores on three AbstRCT medical datasets.S-F1, A-F1, U-F1 and F1 refer to the average macro-F1 score of the support relation, of the attack relation, of no-relation, and of their average, respectively. For DMON the mean results over 5 runs with variance are shown.

Results

Models	Full-F1	F1	R-F1	U-F1
RESARG	-	67.60	38.99	96.20
RESATTARG	-	73.64	50.00	97.28
BERT	32.83	73.89	50.47	97.30
AMCT-Sci	34.58	75.89	54.61	97.17
Roberta	35.96	75.88	54.25	97.51
DMON	**48.37** ± 0.54	**87.36** ± 0.25	**77.03** ± 0.36	**98.70** ± 0.02

Table 2. Experimental results of argument structure learning in terms of macro-F1 scores on the SciDTB datasets. For DMON the mean results over 5 runs with variance are shown. Full-F1, R-F1, U-F1 and F1 refer to the average macro-F1 score of the full label space, the related relation, of no-relation, and of related and no-relation, respectively.

Results

Models	F1	R-F1	U-F1
TSP-PLBA	-	34.00	-
AMPERE++	63.10	-	-
RESATTARG	64.40	30.60	98.30
BERT-Trans	67.80	37.30	98.30
DMON	**68.14** ±0.45	**38.26** ±0.74	**98.37** ±0.06

Table 3. Experimental results of argument structure learning in terms of macro-F1 scores on the CDCP datasets. For DMON the mean results over 5 runs with variance are shown. R-F1, U-F1 and F1 refer to the average macro-F1 score of the related relation, of no-relation, and of their average, respectively.

Ablation study

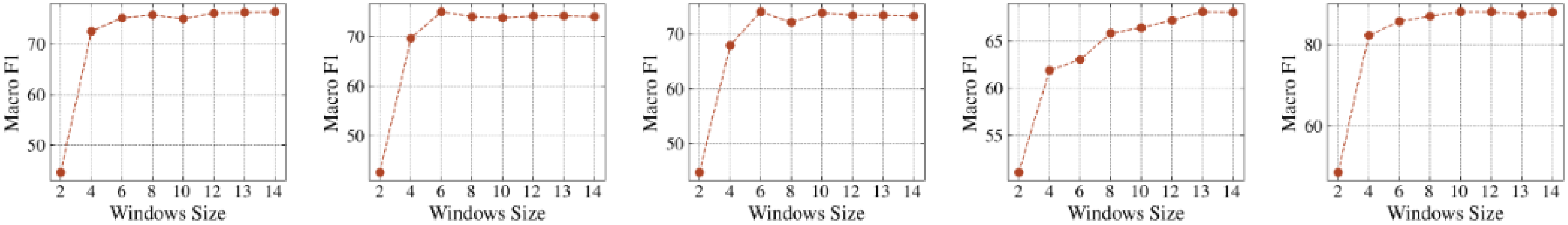
- Bidirectional Learning mechanism
  - Purpose: encode contextual arguments and their relationships to benefit the ASL task
  - Table 4
- Cropping strategy
  - Handle the scarcity of training data
  - Table 5 + Figure 4

Ablation study

Model	Neo	Gla	Mix	CDCP	SCI
DMON	**76.30**	**74.16**	74.07	**68.14**	**87.36**
w/o Voter (h)	74.31	72.04	73.42	67.50	87.08
w/o Voter (t)	75.74	73.83	74.84	67.50	86.58
w/o T	73.74	70.10	71.40	65.39	85.75
w/o H	75.62	71.48	**74.88**	66.55	87.08
w/o H+T	69.10	71.28	70.35	56.24	74.27

Table 4. Experimental results of argument structure learning in terms of average macro-F1 scores on the Neoplasm (Neo), Glaucoma (Gla), Mixed (Mix), and CDCP datasets. w/o Voter (h) or Voter (t) removes the confidence voter and leverages head or tail prediction, respectively. w/o T or H removes either the head branch or tail branch when training the model. w/o H+T completely removes the bidirectional learning mechanism.

Ablation study



(a) Neoplasm Dataset (b) Glaucoma Dataset (c) Mixed Dataset (d) CDCP Dataset (e) SciDTB Dataset

Figure 4. Macro-F1 scores when changing window size of contextual arguments.

Ablation study

Model	Neo	Gla	Mix	CDCP	SCI
DMON$_{\mathrm{win}=13}$	**76.30**	**74.16**	**74.07**	**68.14**	**87.36**
DMON$_{full}$	76.10	73.21	69.82	66.76	87.18

Figure 4. Results of the cropping strategy with window size of 13 (DMON_win=13) and of using the complete relationship tensor (DMON_full) during training.

Conclusion

- A novel framework called the Dual-tower Multi-scale cOnvolution neural Network (DMON) to deal with the ASL task
  - learn the argumentative DAG structure taking into account contextual argumentative relationships.
- SOTA performance on five datasets covering the medical, legal and scientific domains, namely abstRCT-neoplasm, abstRCT-glaucoma, abstRCT-mixed, CDCP, SciDTB.

Reference

- [1] Mayer, Tobias, Elena Cabrio, and Serena Villata. "Transformer-based argument mining for healthcare applications." ECAI 2020. IOS Press, 2020. 2108-2115.
- [2] Stylianou, Nikolaos, and Ioannis Vlahavas. "Transformed: End-to-End transformers for evidence-based medicine and argument mining in medical literature." Journal of Biomedical Informatics 117 (2021): 103767.
- [3] Galassi, Andrea, Marco Lippi, and Paolo Torroni. "Multi-task attentive residual networks for argument mining." IEEE/ACM Transactions on Audio, Speech, and Language Processing (2023).
- [4] Hua, Xinyu, and Lu Wang. "Efficient argument structure extraction with transfer learning and active learning." arXiv preprint arXiv:2204.00707 (2022).
- [5] Niculae, Vlad, Joonsuk Park, and Claire Cardie. "Argument mining with structured SVMs and RNNs." arXiv preprint arXiv:1704.06869 (2017).
- [6] Iskender, Neslihan, et al. "Argument mining in tweets: Comparing crowd and expert annotations for automated claim and evidence detection." International Conference on Applications of Natural Language to Information Systems. Cham: Springer International Publishing, 2021.
- [7] Miller, Tristan, Maria Sukhareva, and Iryna Gurevych. "A streamlined method for sourcing discourse-level argumentation annotations from the crowd." 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). 2019.

Thank you for listening !