IAM: A Comprehensive and Large-Scale Dataset for **I**ntegrated **A**rgument **M**ining Tasks

Liying Cheng[*1,2] **Lidong Bing**[†1] **Ruidan He**[1] **Qian Yu**[‡3] **Yan Zhang**[‡4] **Luo Si**[1]
[1]DAMO Academy, Alibaba Group [2]Singapore University of Technology and Design
[3]The Chinese University of Hong Kong [4] National University of Singapore

Venue : 2022 ACL

발제자 : 이다현

HUMANE Lab

2024-02-20

Motivation

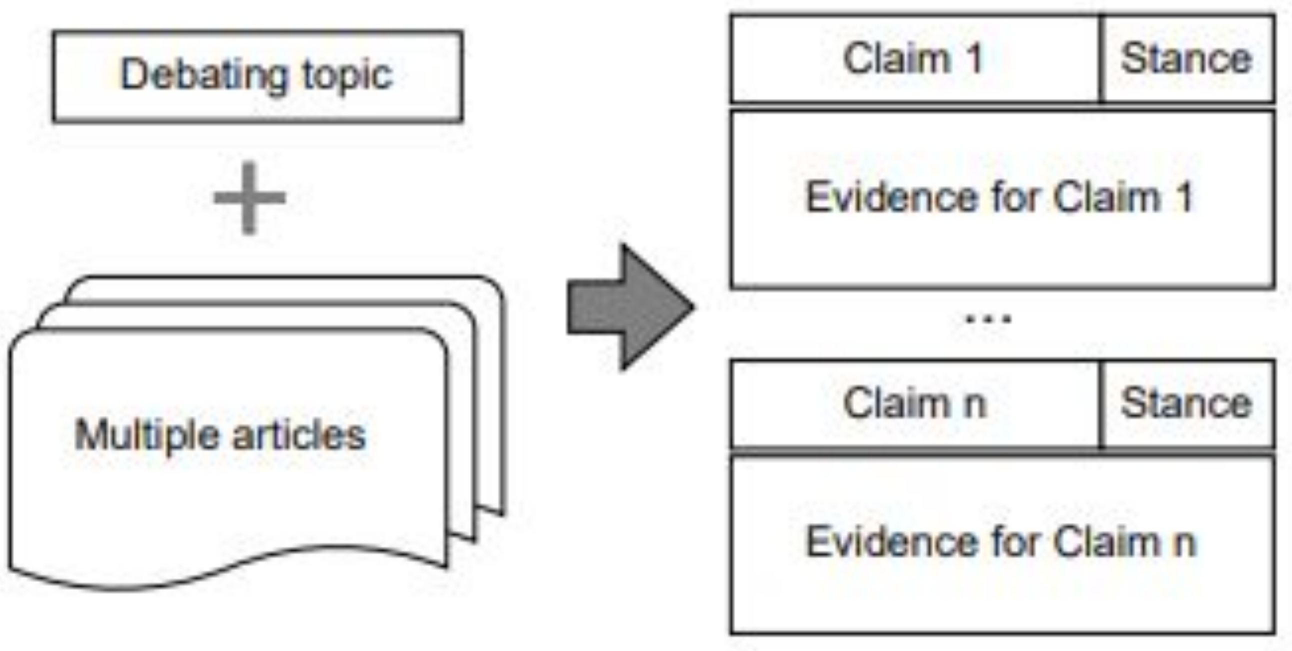


Figure 1: A flow chart showing the debating preparation process.

Motivation

AI타임스

15분 만에 토론 준비 끝... 토론하는 AI 등장

정윤아 기자 입력 2021.03.20 14:53 댓글 0 좋아요 0

IBM, 인간과 토론 가능한 AI기반 시스템 개발
토론 승패보다는 복잡다기한 인간 언어 세계 이해가 목표
아마존 대화형 AI 챌린지 태스크봇(TaskBot) 챌린지 시작

Motivation

Existing argument mining (AM) tasks:

- Claim Extraction
- Stance Classification
- Evidence Extraction

 These essential elements are mutually reinforcing in the debating preparation process

Goal:

- More Integrated tasks

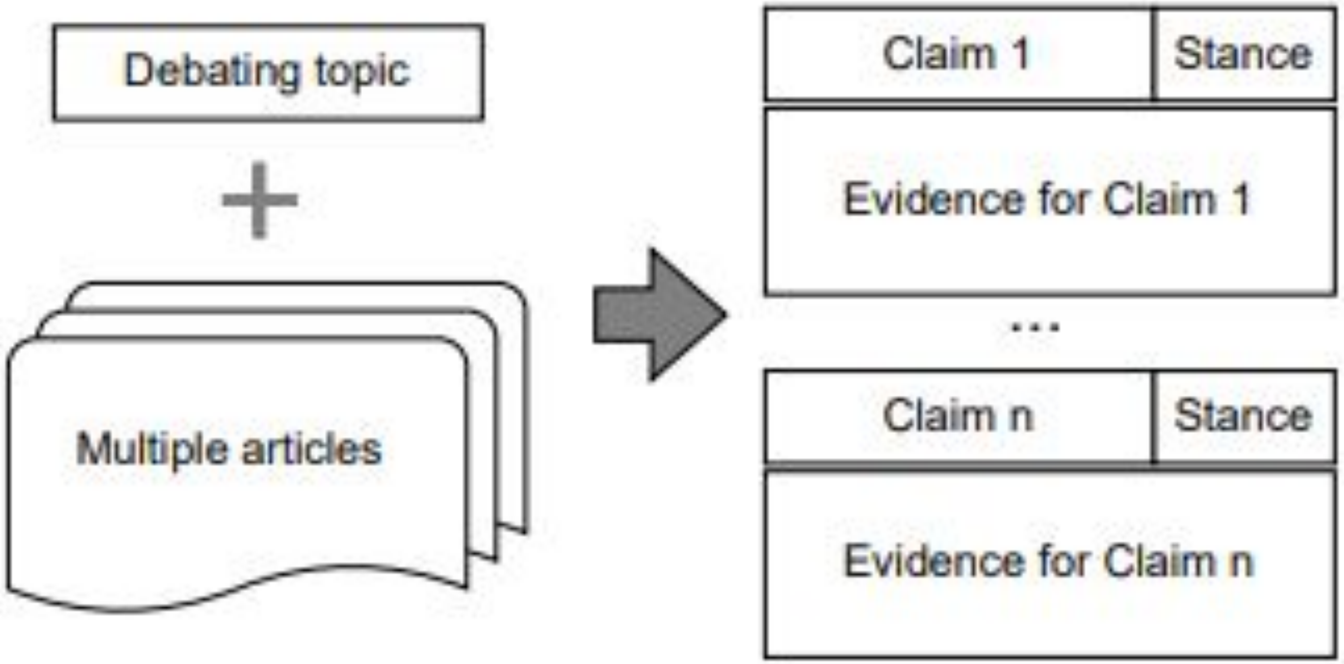


Figure 1: A flow chart showing the debating preparation process.

IAM Dataset – Data Collection

Online forums → 123 debating topics

↓

English Wikipedia → 1010 articles $\xrightarrow{\text{NLTK}}$ 69666 sentences

IAM Dataset – Data Annotation

Claim & Stances → Evidence

<u>A claim</u> is a general and concise statement that directly <u>supports or contests</u> the given topic. (Levy et al., 2014).

	Topic: Will artificial intelligence replace humans	Claim	Stance
1	Job opportunities will grow with the advent of AI; however, some jobs might be lost because AI would replace them.	C_1	+1
2	Any job that involves repetitive tasks is at risk of being replaced.	C_2	+1
3	In 2017, Gartner predicted 500,000 jobs would be created because of AI, but also predicted that up to 900,000 jobs could be lost because of it.		
4	The number of industrial robots has increased significantly since the 2000s.		
5	The low operating costs of robots make them competitive with human workers.		
6	In the finance sector, computer algorithms can execute stock trades much faster than a human, needing only a fraction of a second.		
7	As these technologies become cheaper and more accessible, they will be implemented more widely, and humans might be increasingly replaced by AI.	C_3	+1
8	According to Harvard Business Review, most operations groups adopting RPA have promised their employees that automation would not result in layoffs.	C_4	-1
9	AI is incredibly smart, but it will never match human creativity.	C_5	-1

IAM Dataset – Data Annotation

Claim & Stances → **Evidence**

A piece of evidence is a text segment that directly supports a claim in the context of the topic (Rinott et al. 2015).

	Topic: Will artificial intelligence replace humans	Claim	Stance	Evidence
1	Job opportunities will grow with the advent of AI; however, some jobs might be lost because AI would replace them.	C_1	+1	
2	Any job that involves repetitive tasks is at risk of being replaced.	C_2	+1	
3	In 2017, Gartner predicted 500,000 jobs would be created because of AI, but also predicted that up to 900,000 jobs could be lost because of it.			E_1 | E_2
4	The number of industrial robots has increased significantly since the 2000s.			E_3
5	The low operating costs of robots make them competitive with human workers.			E_3
6	In the finance sector, computer algorithms can execute stock trades much faster than a human, needing only a fraction of a second.			E_3
7	As these technologies become cheaper and more accessible, they will be implemented more widely, and humans might be increasingly replaced by AI.	C_3	+1	
8	According to Harvard Business Review, most operations groups adopting RPA have promised their employees that automation would not result in layoffs.	C_4	-1	E_4
9	AI is incredibly smart, but it will never match human creativity.	C_5	-1	

IAM Dataset – Statics Comparison

	Topics	Articles	Articles with claims	Claims	Support	Contest	Claims with evidence	Evidence	CEPs
Levy et al. (2014)	32	326	-	976	-	-	-	-	-
Rinott et al. (2015)	39	274	274	1,734	-	-	1,040	3,057	5,029*
Aharoni et al. (2014)	33 (12)	586	321 (104)	1,392	-	-	(350)	(1,291)	1,476*
Bar-Haim et al. (2017)	55	-	-	2,394	1,324	1,070	-	-	-
IAM (Ours)	123	1,010	814	4,890	2,613	2,277	3,302	9,384	10,635

Table 2: Overall statistics comparison of the existing datasets and our dataset. Note that: (1) in Aharoni et al. (2014)'s dataset, the numbers in the parenthesis refer to the evidence labeling data; (2) the numbers with * are calculated by us since they are not shown in the original papers.

Comprehensive & Large-Scale

Existing Tasks

Existing tasks:

- Task 1 – Claim Extraction
- Task 2 – Evidence extraction
- Task 3 – Stance classification

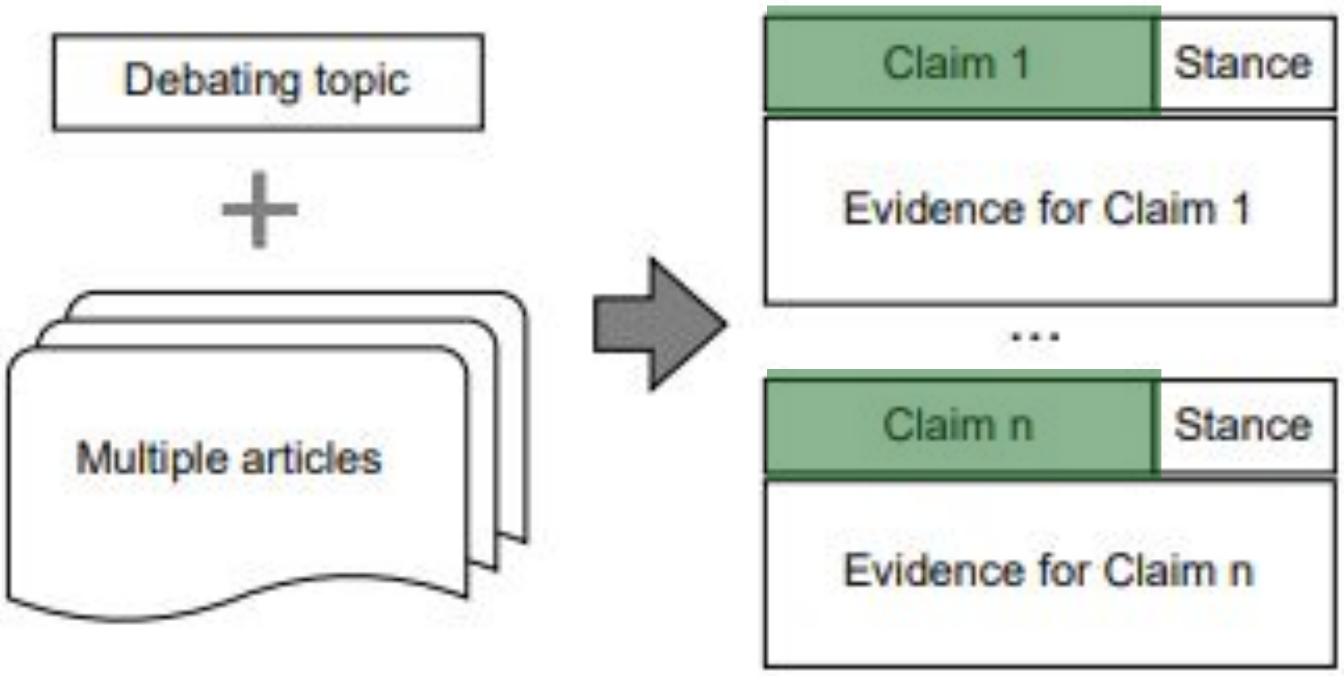


Figure 1: A flow chart showing the debating preparation process.

Existing Tasks

Existing tasks:
- Task 1 – Claim Extraction
- **Task 2 – Stance classification**
- Task 3 – Evidence extraction

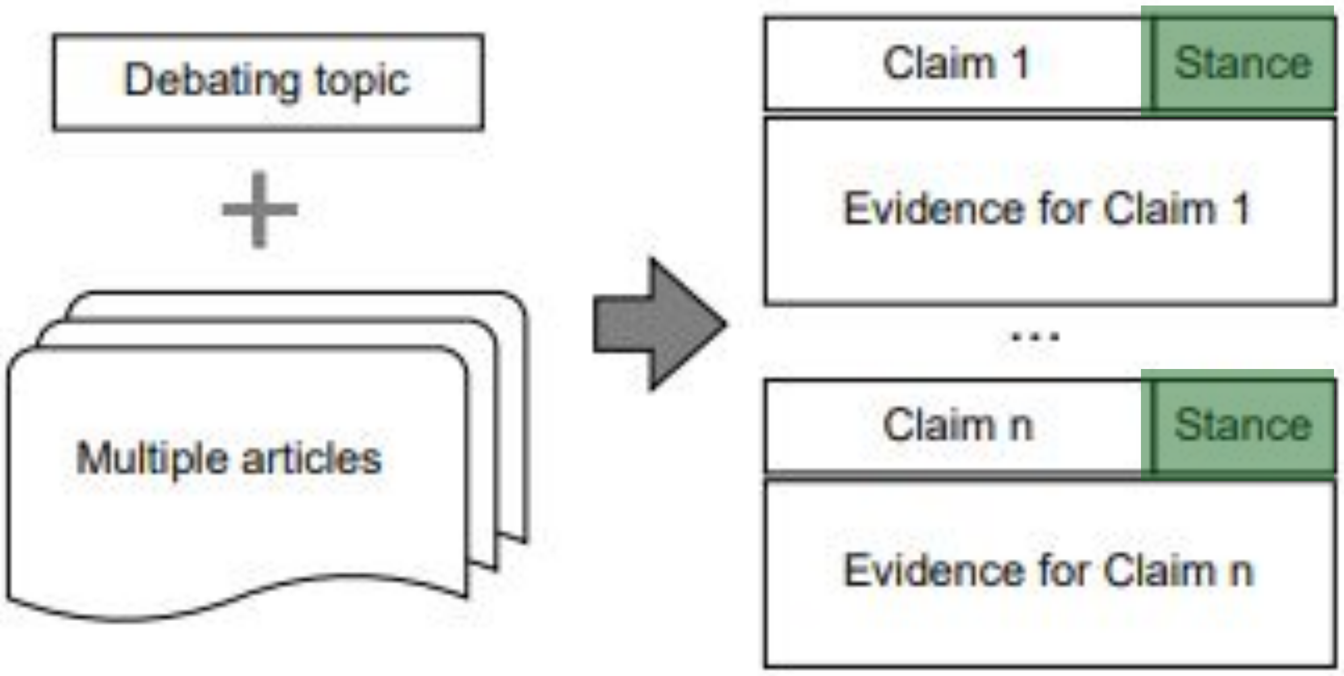


Figure 1: A flow chart showing the debating preparation process.

Existing Tasks

Existing tasks:
- Task 1 – Claim Extraction
- Task 2 – Stance classification
- Task 3 – Evidence extraction

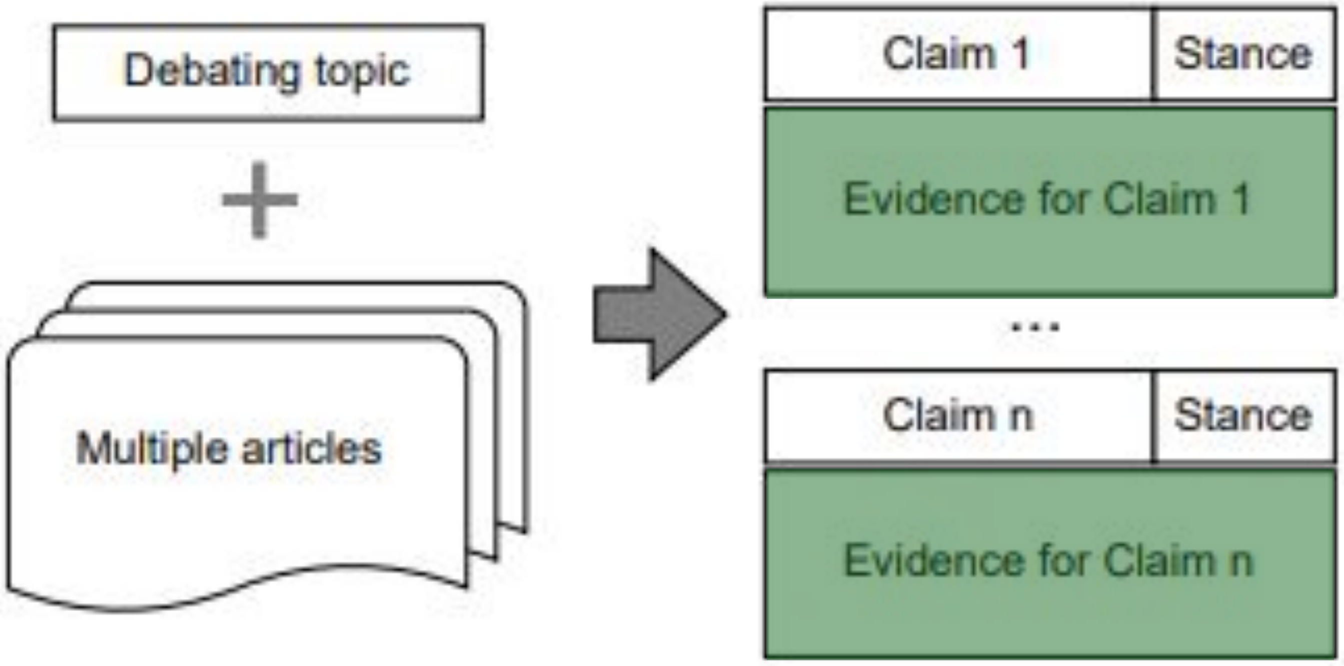


Figure 1: A flow chart showing the debating preparation process.

New Integrated Tasks

Existing tasks:
- Task 1 – Claim Extraction
- Task 2 – Stance classification
- Task 3 – Evidence extraction

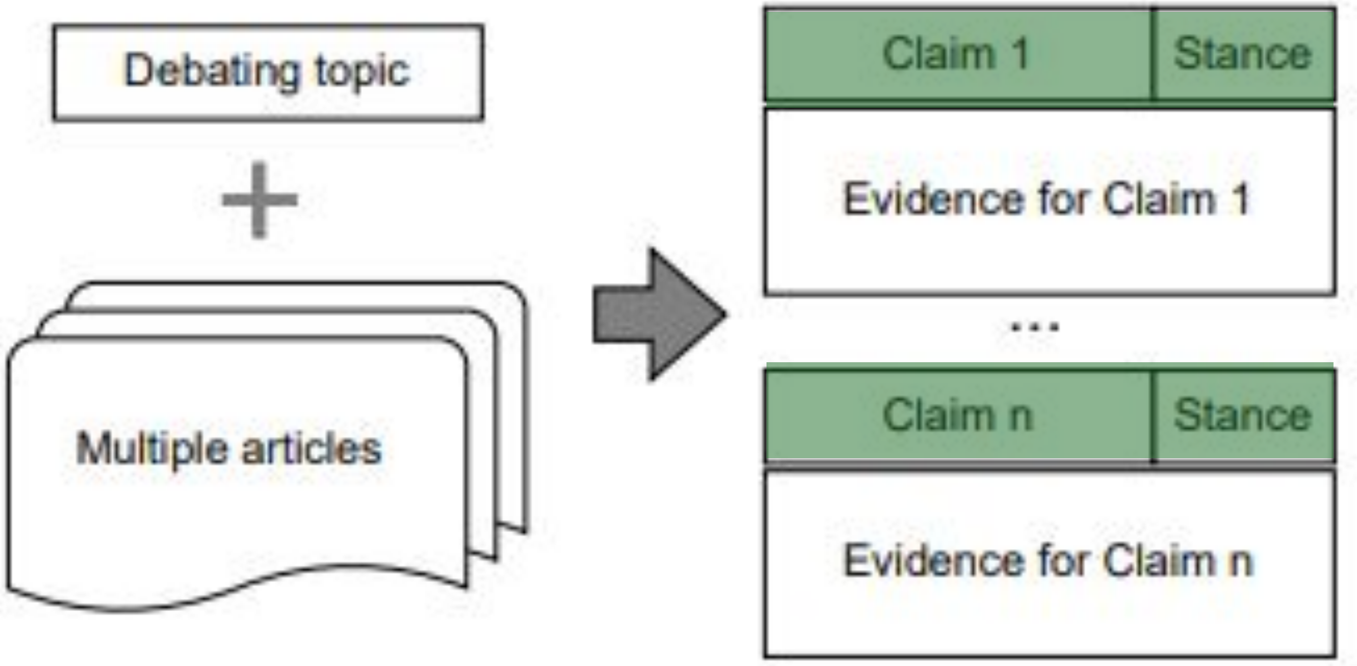


Figure 1: A flow chart showing the debating preparation process.

New integrated tasks:

- Task 4 – Claim extraction with stance classification(CESC)
- Claim-evidence pair extraction (CEPE)

New Integrated Tasks

Existing tasks:
- Task 1 – Claim Extraction
- Task 2 – Stance classification
- Task 3 – Evidence extraction

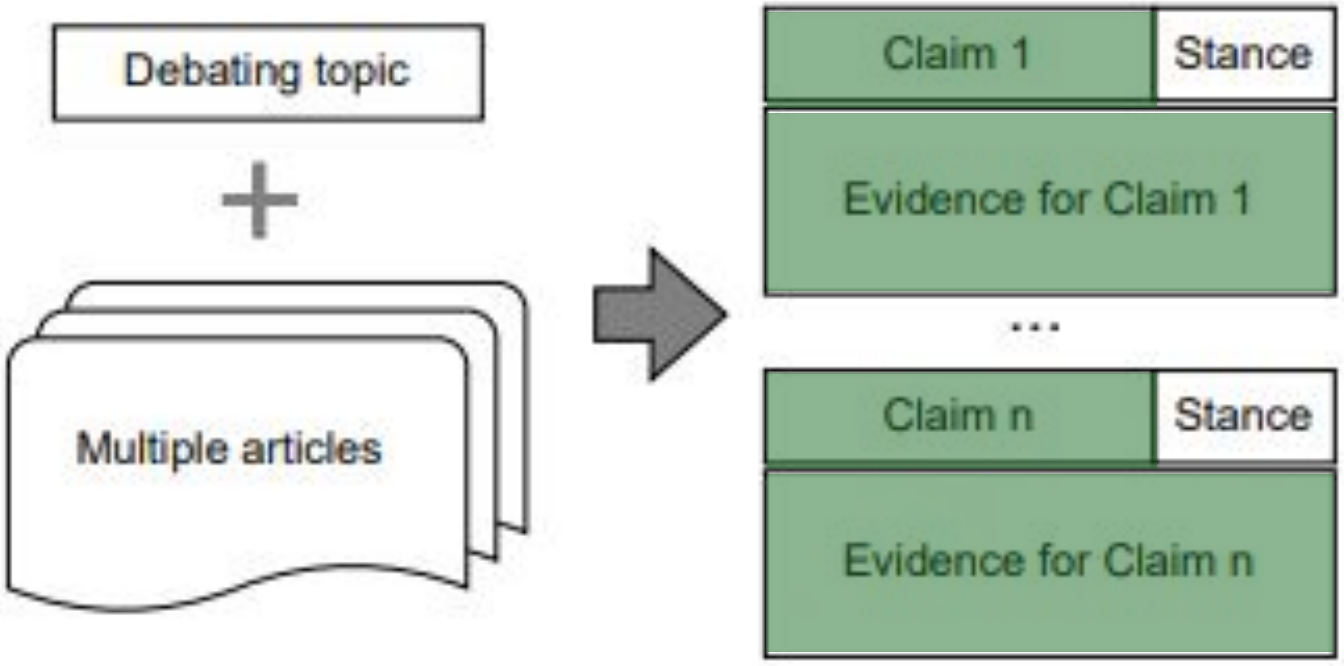


Figure 1: A flow chart showing the debating preparation process.

New integrated tasks:

- Task 4 – Claim extraction with stance classification(CESC)
- Task 5 – Claim-evidence pair extraction (CEPE)

Approaches

Existing tasks:

- Task 1 – Claim Extraction
- Task 2 – Stance classification
- Task 3 – Evidence extraction

Sentence pair

New integrated tasks:

- Task 4 – Claim extraction with stance classification(CESC)

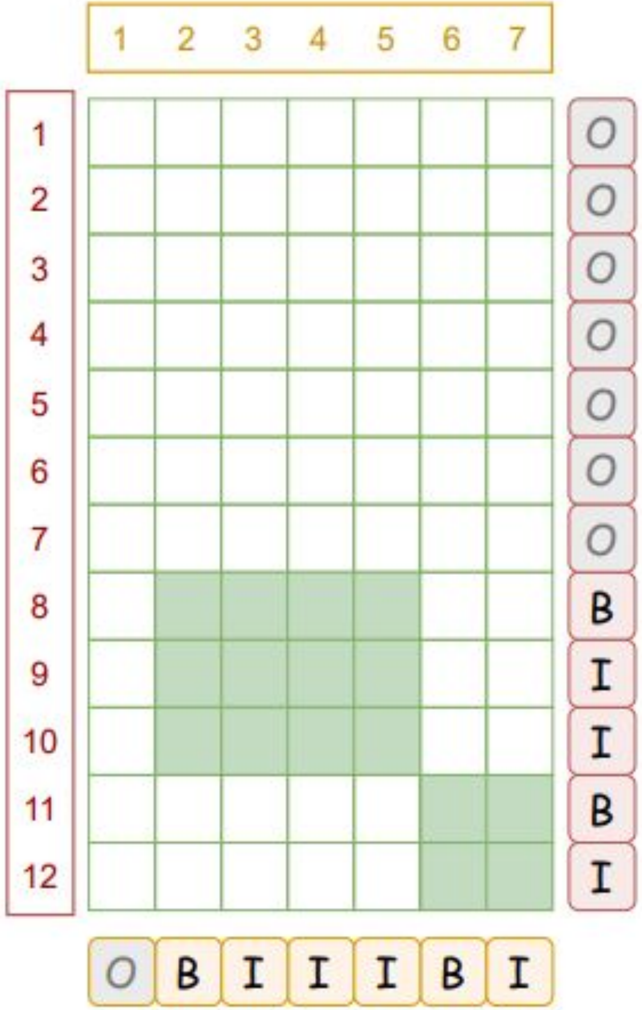
Pipeline: Task1 +

Multi-Label

- **Task 5 – Claim-evidence pair extraction (CEPE)**

Pipeline: Task1 +

Multi-Task model (Cheng et al.



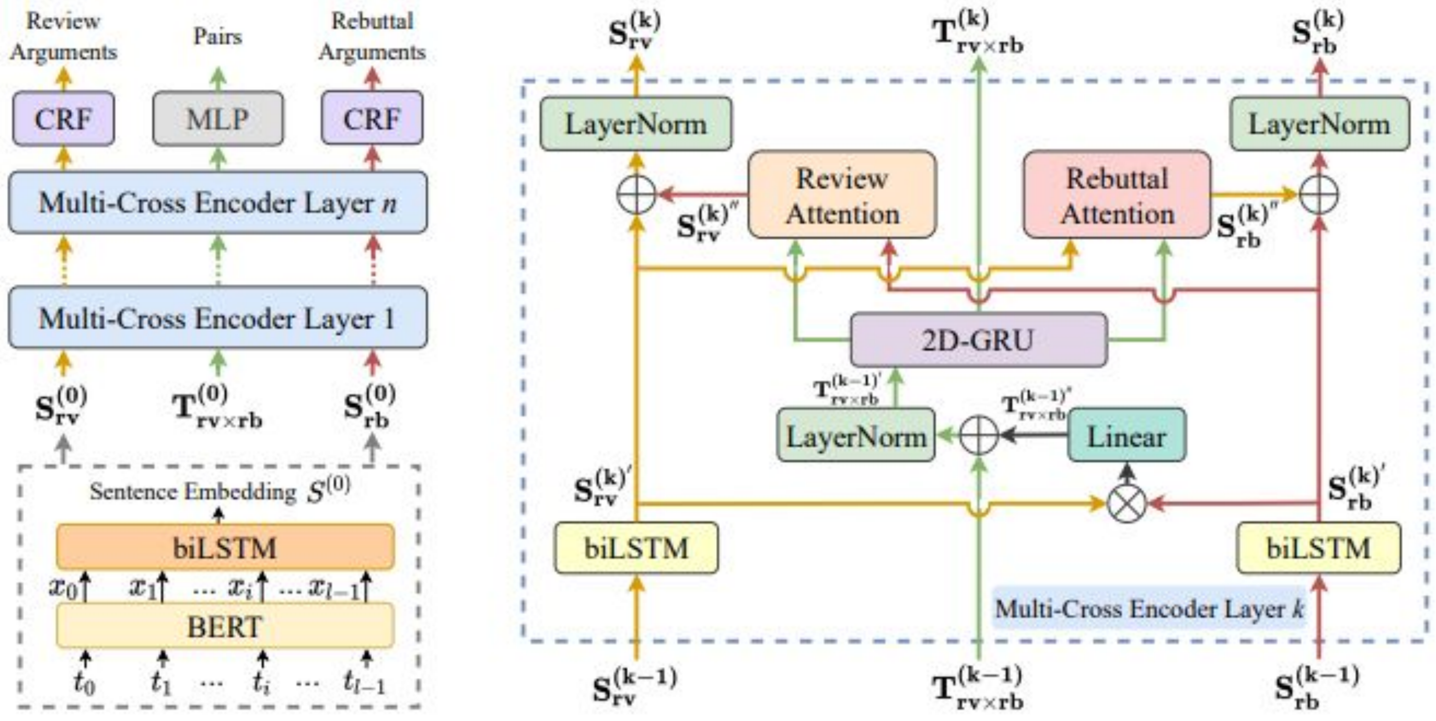


Figure 2: Overview of our model architecture with n multi-cross encoder layers (shown on the left). The sentence embedder (in the grey dotted box at the bottom left) shows the process of obtaining initial review and rebuttal sentence embeddings from pre-trained BERT token embeddings with biLSTM. The kth multi-cross encoder layer (in the blue dotted box on the right) shows the process of getting the sentence representations of review and rebuttal and the pair representations for the next layer.

the attention-guided multi-cross encoding-based model

- Task 5 – Claim-evidence pair extraction (CEPE)

Multi-Task model (Cheng et al.

Experiments – Claim Extraction

Models	Macro F_1	Micro F_1	Claim F_1
BERT-base-cased	72.08	**92.51**	48.08
RoBERTa-base	**72.36**	91.09	**50.35**

Table 5: Claim extraction performance.

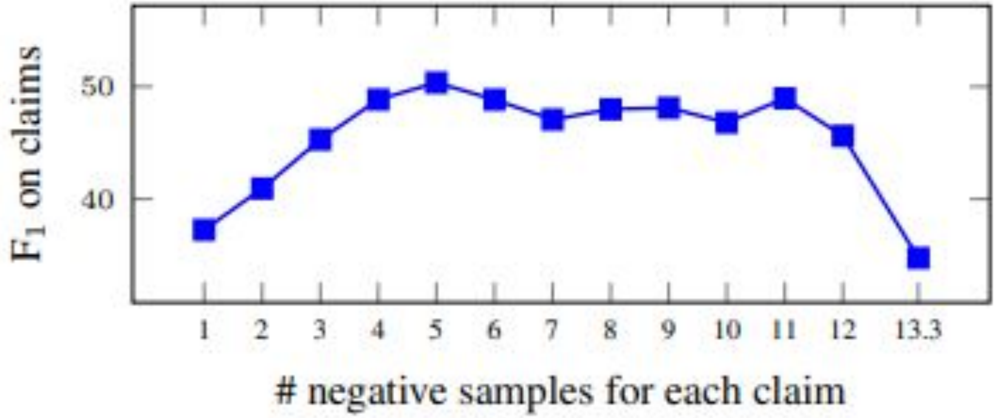


Figure 2: Effect of negative sampling for claim extraction with RoBERTa-base model.

Experiments – Stance Classification

Models	Acc.	Support F_1	Contest F_1
BERT-base-cased	73.43	73.08	73.78
RoBERTa-base	**81.21**	**81.21**	**81.21**

Table 6: Stance classification performance.

Experiments – Evidence Extraction

Models	Macro F_1	Micro F_1	Evi. F_1
BERT-base-cased (T+C)	58.17	72.75	38.15
RoBERTa-base (T+C)	62.43	78.13	**40.89**
BERT-base-cased (C)	58.01	72.65	37.92
RoBERTa-base (C)	**63.37**	**80.29**	40.16

Table 7: Evidence extraction performance.

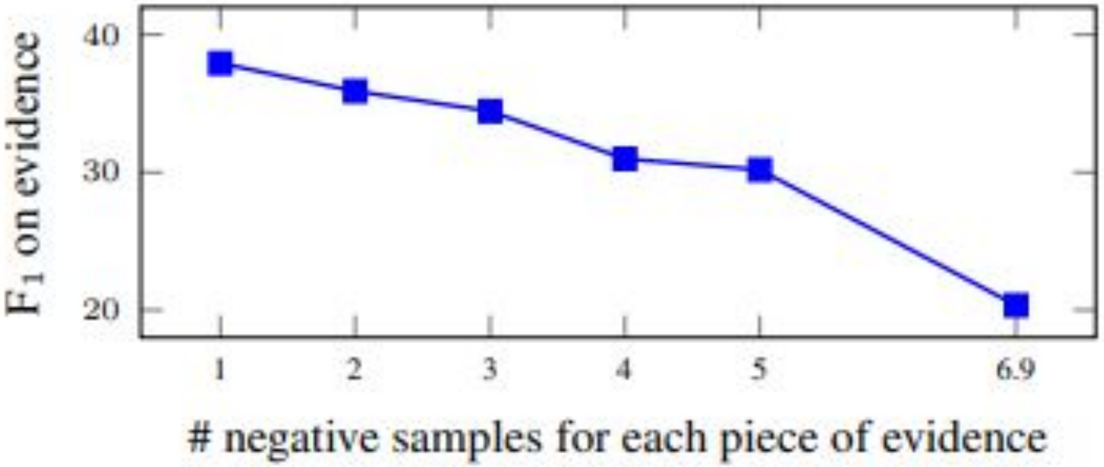


Figure 3: Effect of negative sampling for evidence extraction with BERT-base-cased (C).

Experiments – Claim Extraction with Stance Classification(CESE)

Models	Macro F_1	Micro F_1	Support F_1	Contest F_1
Pipeline	55.95	88.56	33.39	40.10
Multi-label	**60.25**	**91.22**	**38.34**	**47.31**

Table 8: CESC task performance.

Experiments – Claim-evidence pair extraction (CEPE)

Models	Precision	Recall	F_1
Pipeline	16.58	22.11	18.95
Traversal	24.06	**38.74**	29.69
Multi-task	**43.54**	30.57	**35.92**

Table 9: CEPE task performance.

Conclusion

- Argument Mining Task들을 위한 데이터들을 통합하고 주석처리한 대규모 데이터셋 IAM 구축
- 실용적이고 효과적인 통합 Task인 CESC와 CEPE를 처음으로 제안
- 새로운 데이터셋으로 지금까지의 Argument mining task들에 대한 실험을 수행

Open question

- LLM을 활용해서 비슷하게 주석 처리된 Argument Mining 데이터셋을 만들 수 있을까?