

판결문 논증 분석을 위한 트랜스포머 모델 기반 논증 구조 추출 방안에 관한 연구*

구예리¹ 박성미² 안정민¹

¹ 한림대학교 정보법과학전공

² 고려대학교 정보보호대학원

yerie.gu@gmail.com, outspark@korea.ac.kr, jjia@hallym.ac.kr

A Transformer-based Method for Extracting Argument Structures in Court Decisions: An Empirical Study

Yeri Gu¹ Sungmi Park² Jungmihn Ahn¹

¹Legal Informatics and Forensic Science Major, Hallym University

²School of Cybersecurity, Korea University

요 약

최근 인공지능 기술의 발전으로 법률 및 수사 분야에서 자연어 처리 기술 활용이 증가하고 있다. 수사 분야에서는 형사소송법이 개정됨에 따라 증거 기반의 논리적 논증이 중요시되고, 이에 객관적이고 균일한 수사 품질의 보장을 위한 논증 기반 사건 검증 시스템 필요성 대두되는 바이다. 따라서 본 논문에서는 법률 분야에서의 인공지능 기술 활용 및 발전 가능성을 탐구하기 위해 트랜스포머 모델을 활용한 법적 논증 구조 자동 추출 시스템을 제안한다.

1. 서 론

최근 인공지능 기술의 발전으로 자연어처리 기술을 활용하여 법률 및 수사 분야의 다양한 문제들을 해결하려는 시도가 활발히 이루어지고 있다. 경찰 수사 분야의 경우, 개정된 형사소송법 및 검찰청법의 시행으로 경찰의 수사 책임이 강조됨에 따라 객관적 증거를 기반으로 한 논리적 사실 주장이 더욱 요구되고 있다. 그러나 기존의 사건 분석 지원 도구는 증거 수집 및 분석에 중점을 두고 있어, 논리적인 검증에는 한계가 존재한다. 이에 본 연구는 자연어 처리 방법 중 텍스트에서 논증을 식별하고 분석하는 것을 목적으로 하는 논증 마이닝(Argument Mining) 기법을 통해 입력된 수사문서에서 논증을 자동 추출하고, 추출된 논증 요소들 간의 관계를 자동 분류하여 논증 구조 그래프를 제공함으로써 수사 관계자들이 빠르고 객관적인 시각으로 사건의 논증 구조를 검토할 수 있는 모델 구조를 개발하는 것을 목적으로 한다.

이를 위해 본 논문에서는 현행 수사결과보고서와 유사한 구조를 가지는 제1심 형사판결문 256건을 대상으로 툴민(Toulmin)의 논증구조를 본 연구의 목적에 맞게 확장 및 재개념한 툴민+ 모델을 기반으로 논증 데이터셋을 생성하고, 이러한 일련의 과정에 최근 자연어 처리 분야에서 활발히 사용되고 있는 트랜스

포머(Transformer) 기반 사전학습 언어모델을 사용하여 그 성능을 높이고자 한다. 이를 통해 본 연구는 수사 검증을 위해 필수적인 논증 분석 시스템을 개발함으로써 경찰에 요구되는 핵심 역량을 제고하고 사법체계의 효율성을 높이는 데 기여하고자 한다.

2. 관련 연구

2.1. 툴민(Toulmin) 논증 구조

툴민의 논증 구조는 체계적이고 구조화된 방식으로 법적 논증을 분석하기 위해 고안된 논증 모델로, 6개의 논증 요소(Datum, Claim, Warrant, Backing, Qualifier, Rebuttal)를 기반으로 법적 추론의 결여점을 파악하고 증거의 강도를 평가한다 [1].

2.2 논증 마이닝

자연어 텍스트로부터 논증 구조를 추출하는 논증 마이닝 시스템은 일반적으로 논증 요소 식별과 논증 구조 예측의 두 단계 파이프라인으로 구성된다 [2]. 첫번째 태스크의 경우, 주어진 문장에 대해 적용된 논증 구조에 해당하는 구성요소(예: 주장, 전제)를 분류하는 문제로 SVM(Support Vector Machines), Naive Bayes, Logistic 회귀 등과 같은 전통적인 기계학습 방법들이 제안된 바 있다 [3]. 다음으로는 식별된 논증 요소 간의 관계(예: 공격, 지지)를 예측하는 태스크로, 높은 수준의 추론 능력이 요구되며 파악된 논증 관계를 바탕으로 논증 그래프를 구성한다.

* 이 논문은 2021년도 정부(경찰청)의 재원으로 지원받아 수행된 연구결과임 [내역사업명: AI 기반 범죄수사 지원/연구개발과제번호: PR10-02-000-21]

3. 실험

3.1 툴민+ 논증 데이터셋

본 연구에서는 판결문 논증 구조를 추출하기 위해 기존 툴민의 논증 모델을 본 연구의 목적에 맞게 수정 및 확장한 툴민+ 모델을 <표 1>과 같이 새롭게 정의하고, 이를 바탕으로 1심 형사 판결문 256건을 주석하여 데이터셋을 구축하였다. 툴민+ 논증 모델은 총 7개의 논증 구성요소를 가지며, 각 요소들 사이의 관계는 공격이나 지지 등의 관계 구성 요소를 사용하여 표현된다. 툴민+구조는 추가된 요소와 관계 표현을 통해 판결문 내 대립되는 변호인과 판사의 논증구조를 명확히 표현할 수 있다. 최종 구성된 데이터셋은 그림 1과 같이 판결문 내 자연어 구절에 대해 태깅된 논증 요소, 관련 있는 논증 요소, 요소간의 관계를 어노테이션 한 결과를 포함한다.

논증 요소	설명
Datum(D)	주장(claim)을 뒷받침하는 사실이나 증거
Expert Opinion (EO)	전문가가 제공하는 증언 또는 문서
Backing (B)	판례나 법률에 대한 참조
Warrant (W)	판례나 공통 지식
Inference (I)	증거(datum)와 주장(claim) 사이의 추론
Claim (C)	판사와 변호사가 제시하는 쟁점
Issue Conclusion (IC)	쟁점의 결론

표 2 툴민+ 논증 모델 요소 설명[4]

```
{ "annot_data": [
  { "toulmin_no": 1,
    "component": "1_d_1",
    "relation": "1_c_1",
    "relation_type": "s",
    "phrase": "피고인이 피해자의 허벅지를 칼로 1회 찔러 상해를 가한 사실은 있으나"
  }
]}
```

그림 1 툴민+ 논증 데이터셋 예시

3.2. 모델 설계

본 연구에서 제안하는 모델은 트랜스포머 기반 언어 모델을 사용하여 판결문 내 논증 구조를 식별하는 것을 목표로 한다. 본 모델은 그림 2와 같이 다중 모듈로 구성되어 있으며, 순차적으로 논증 구성요소를 식별한 다음, 논증적으로 관련된 구문 쌍과 그들의 관계 유형을 분류하고, 최종적으로 논증 구조를 추출하고 시각화하는 구조를 가진다.

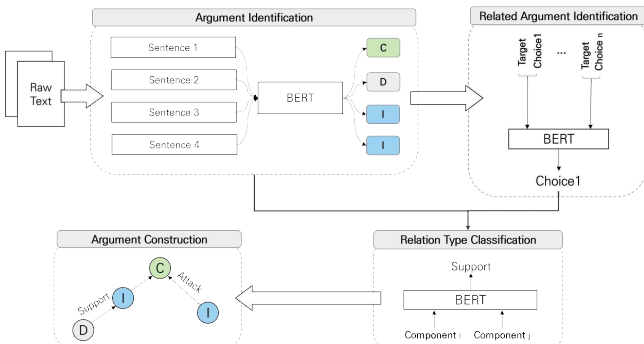


그림 2 제안하는 논증 구조 추출 모델 구조

3.2.1. 논증 요소 분류

논증 구조 추출 모델 구조의 첫번째 단계인 논증 요소 분류 모듈의 목표는 말뭉치의 각 구문에 대해 라벨링된 논증 요소를 올바르게 분류하는 것으로, 본 연구에서는 해당 태스크를 다중 클래스 분류 문제로 정의하고 사전 학습된 한국어 BERT 모델인 KoBERT를 fine-tuning 하는 방식으로 접근한다.

3.2.2. 논증 관계 분류

다음으로는 분류된 논증 구성 요소들을 대상으로 요소들간의 관계를 탐지를 진행한다. 논증 관계를 인식하는 것은 (1) 텍스트로부터 관련된 논증 쌍을 식별하고 (2) 해당 논증 쌍의 관계 유형 분류하는 두가지의 상호 관련된 절차로 진행되므로 본 연구에서는 관계 분류를 위한 두가지의 별도 모듈을 구성하였다.

3.2.2.1. 논증 관계 후보 분류

논증 관계 분류 모델의 첫번째 태스크는 가능한 논증 쌍을 식별하는 것으로 본 연구에서는 이 문제를 타겟 구절에 대해 여러가지 논증 조합을 생성하고, 그 중 가장 논증적으로 관련성이 높은 쌍을 예측하는 Multiple-Choice 문제로 정의한다. 이 접근 방식은 모델이 주어진 선택지에서 문맥적 관계를 학습하게 되므로 모델의 추론 능력이 단순 문장-대-문장의 관계를 학습하는 것보다 상당히 향상될 것이라고 가정한다. 모델은 사전 학습된 KoBERT 모델의 경량화 된 버전인 DistilKoBERT를 사용하였으며, 그림 3과 같이 타겟 구성요소와 관계 후보 중 하나의 선택지를 시퀀스로 연결하여 4개의 선택지 중 올바른 답을 선택하도록 학습되는 구조를 가진다. 각 시퀀스는 최종적으로 Softmax 레이어를 통해 가장 높은 확률을 가지는 시퀀스를 분류한다.

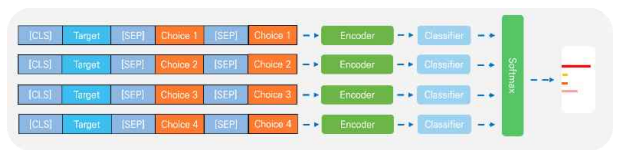


그림 3 BERT 기반 Multiple-Choice 분류 모델 구조

3.2.2.2. 논증 관계 분류

논증 관계 쌍을 식별한 후, 논증 구조를 추출하는 태스크를 수행하기 위해 이를 두 개의 텍스트 입력에 가장 가능성이 높은 클래스를 할당하는 시퀀스 분류 문제로 정의하고, NLI(Natural Language Inference) 기반의 접근 방식을 사용하여 문장 간의 관계를 추론하였다. 이를 위해 NLI를 포함한 8개의 downstream 태스크를 수행하는 한국어 사전학습 모델인 KLUE-BERT 모델을 fine-tuning하여 주어진 논증 쌍의 관계가 지지, 공격, 또는 병렬인지 예측하도록 구성하였으며 모델의 구조는 그림 4와 같다.

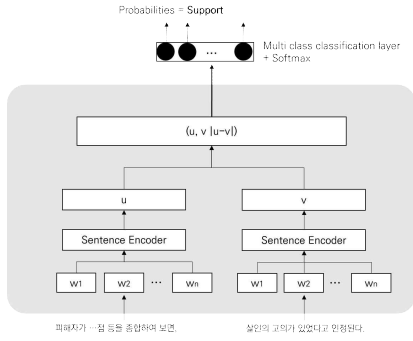


그림 4 NLI 모델 구조

4. 결과

4.1. 모델 성능 평가

제한한 모델의 세가지 모듈에 대한 성능 측정을 위해 F1-점수, 정밀도, 재현율을 함께 도출하였으며 그 결과는 <표 2>와 같다.

표 2 판결문 논증 구조 추출 모델별 성능 평가

태스크		F1	Precision	Recall
논증 요소 분류 (KoBERT)		0.92	0.92	0.92
계논증 분류관	논증 관계 후보 분류 (DistiBERT)	0.75	0.76	0.75
	논증 관계성 분류 (KLUE-BERT)	0.91	0.91	0.91

<표 2>의 결과를 살펴보면 세가지 태스크 모두에서 제안한 모델이 70% 이상의 F1-점수가 계산됨을 확인할 수 있었고, 논증 요소 분류와 관계성 분류의 경우 90% 이상의 우수한 성능을 보임을 확인할 수 있었다.

논증 관계 식별의 경우, 논증의 연결성과 관계적 속성을 이해하기 위해 고수준의 지식 표현 및 추론 능력을 요하므로 관계 요소별 분류 성능을 살펴보았다. 대부분의 레이블에서 모델은 좋은 성능을 보였지만, 공격 레이블 분류에 대해서는 성능이 80%로 하락하였는데, 이에 대하여 데이터셋 내 클래스 불균형이 모델 성능 하락의 영향을 미쳤다고 추론할 수 있다.

표 3 논증 관계 요소별 분류 성능

레이블	실험셋	Precision	Recall	F1-score
Attack	159	0.8000	0.8054	0.8027
No-relation	297	0.9959	1.0000	0.9979
Parallel	1,094	0.9182	0.9525	0.9351
Support	1,012	0.9459	0.9061	0.9256

마지막으로 자동 분류된 논증 요소와 논증 관계 정보를 매핑하여 논증 그래프를 추출한 결과, 총 512개의 논증 구조를 추출하였고 해당 구조들이 공식화된 툴민+ 논증 구조와 부합함을 확인할 수 있었다.

표 4 판결문에서 자동 추출된 논증 구조 그래프 예시

텍스트	추출 그래프
사건명: 수원지방법원2014고합317	
<p>W: 현주건조물방화치사죄 ... 고의가 있어야 하고, I: 이는 ... 할 것이며, I: 이 경우 ... 할것인데, D: 앞서 본 바와 같이 ...하였는바, I: 당시 ...하였거나, C: 적어도 ... 하므로, IC: 피고인 및 변호인의 ...받아들일 수 없다.</p>	

5. 결론

본 연구에서는 범죄 수사 시 논리적 사건 검토 과정에 도움이 될 수 있는 논증 구조 추출 시스템을 제안하였다. 이를 위해 한국어 판결문 논증 구조 분석에 적합한 툴민+ 논증 모델을 개발하고, 이를 기반으로 논증 구성 요소 및 관계를 태깅한 새로운 말뭉치를 구축하였다. 구축된 데이터셋을 대상으로 한국어 사전 학습 BERT 모델을 사용하여 논증 구성 요소를 분류하였으며, 논증 관계 분류 작업에 대해서는 두 단계의 세부 모듈을 구성하여 Multiple-choice 모델로 논증 관계 쌍을 찾고, BERT 기반 NLI 모델을 사용하여 논증의 관계적 성질을 추가로 분류하였다. 실험 결과, 본 연구에서 제안하는 트랜스포머 기반 구조를 사용하여 판결문 텍스트로부터 논증 구조를 성공적으로 추출할 수 있음을 확인하였다. 또한 자동 추출된 논증 구조가 수동으로 분석한 판결문 내 논증 패턴과 부합한다는 것을 확인하여 본 연구에서 제안하는 시스템이 판결문의 논증 구조를 성공적으로 식별할 수 있다는 것을 증명하였다.

참고 문헌

- [1] S. Toulmin, *The uses of argument*, Updated ed. Cambridge, U.K. ; New York: Cambridge University Press, 2003.
- [2] G. Zhang, P. Nulty, and D. Lillis, "A Decade of Legal Argumentation Mining: Datasets and Approaches," in *Natural Language Processing and Information Systems*, vol. 13286. Cham: Springer International Publishing, 2022, pp. 240-252.
- [3] E. Cabrio and S. Villata, "Five Years of Argument Mining: a Data-driven Analysis," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization, Jul. 2018, pp. 5427-5433.
- [4] 한림대학교, 정보법과학연구소, "논증구조 수사데이터분석 가이드라인V2.0", 2022.