

## 동기 기반 분류 데이터셋을 활용한 유사 사건 검색 연구:

## 형사 판결문을 중심으로\*

최아리<sup>01</sup> 박성미<sup>2</sup> 박노섭<sup>1</sup><sup>1</sup>한림대학교 정보법학과<sup>2</sup>고려대학교 정보보호대학원

arichoi@hallym.ac.kr, outspark@korea.ac.kr, rspark@hallym.ac.kr

## Motive-based Classification Dataset Research for Similar Case Search

## Focusing on Criminal on Judgments

Ari Choi<sup>01</sup> Sung-Mi Park<sup>2</sup> Ro-seop Park<sup>1</sup><sup>1</sup>Legal Informatics and Forensic Science, Hallym University<sup>2</sup>School of Cybersecurity, Korea University

## 요약

인공지능 기술의 발전으로 리걸테크 분야에서 대용량의 법률 문서를 효율적으로 활용하기 위한 방안들이 연구가 되고 있으며, 판결문을 대상으로 유사 사건을 검색하는 서비스가 등장하였다. 본 연구에서는 형사 하급심 판결문을 대상으로 효율적인 유사 사건 검색 시스템 개발하기 위해 임베딩 모델의 결과를 비교하고, 수사 전문가들의 평가를 통해 유사 사건 판단 기준을 연구하였다. 이를 통해 여러 유사 기준 중 주로 사용된 판단 기준이었던 동기에 중점을 두어 유사성을 판단할 수 있는 판결문 분류 데이터셋을 구축하였고, Random Forest, SVM, Decision Tree, KoBERT, KLUE(bert-base) 모델을 사용하여 그 유효성을 검증하였다

## 1. 서론

4차 산업혁명의 도래로 인공지능 기술을 활용하면서 처리해야 할 데이터가 기하급수적으로 증가하였고, 리걸테크 분야에서도 인공지능을 활용하여 대용량의 법률 문서를 효율적으로 활용하기 위한 방안들에 대한 연구가 늘어나고 있다. 특히 법령 또는 판례 검색의 경우, 단순히 키워드를 검색하는 수준을 넘어 기준 사건과 비슷한 사건을 도출하는 유사 사건 검색까지 가능해지고 있다. 유사 사건 검색은 사용자들에게 빠르고 정확한 유사 사건을 도출하여 효율적인 검색 기능을 제공하는 데 그치는 것이 아니라 유사 사건 검색 결과로 사건 당사자들의 의사결정에 큰 도움을 줄 수 있다. 우리나라 내 이러한 유사 사건 검색 기능을 제공하고 있는 곳은 엘박스(LBox)와 빅케이스(Big Case) 등이 있다. 엘박스는 범죄사실 단위 또는 사용자가 원하는 문단을 지정하여 유사 사건 검색을 할 수 있는 기능을 제공하고, 빅케이스에서는 판결문 전체를 대상으로 유사 사건을 검색할 수 있다. 하지만, 두 곳 모두 유사 사건에 대한 기준을 공개하지 않고 설명도 제공하고 있지 않아 '유사 사건'을 명확히 정의하고 있다고 보기 어렵다.

이러한 내용적 유사성 외에도 유사성을 비교할 심급을

선정하는 것 역시 중요하다. 우리나라는 3심 제도를 채택하고 있고 법원에 따라 다루는 법적 쟁점이 달라진다. 이에 법률 전문가들은 인공지능을 활용해 판결문을 분석하고 그 결과를 기반으로 유사 사건 검색이나 다른 태스크들을 진행하기 위해서는 직접적인 사실관계를 다루는 하급심이 필요하다며 하급심의 중요성을 강조하였다[1]. 따라서 본 논문에서는 형사 하급심 판결문을 대상으로 유사 사건 검색을 진행하고, 형사 판결문을 위한 유사 사건의 기준을 수립하였다. 이후, 해당 기준을 활용한 데이터셋 구축을 진행하고 다양한 모델을 통해 데이터셋을 검증하였다.

## 2. 형사 판결문 대상으로 유사사건 검색 실험 및 결과

유사 사건 검색에는 크게 두 가지 방법을 사용한다. 먼저 텍스트 기반 의미적 유사도 (Semantic Similarity)를 사용하는 방식이 있으며, [2]는 교통사고법과 관련된 판결문을 수집해 SBERT를 이용해 유사도를 측정하고, 유사 판결문 데이터셋을 구축하였다. 하지만 해당 논문에서는 '유사성' 판단 기준에 대해 별도로 언급하지 않았다.

또 다른 방식으로는 유사도를 측정하고자 하는 대상 문서의 특성을 반영해 유사도를 측정하는 방식이다. 하지만 '유사성'의 기준은 객관적으로 확립되어 있지 않아, 수사관의 경력 등 개인의 판단에 따라 달라질 수 있다. [3]에서는 특정 단어의 빈도수를 확인하거나 어휘의 의미적 관계를 나무형태로 분류하는 방법을 통해 유사성을 판단하는 것은 범죄와 관련된 특정 목적을 가진 텍스트

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

에는 적합하지 않을 수 있어 침입 절도 성격에 맞춰 주요 수법 항목과 부가 수법 항목으로 나누었고, 문서에서 추출한 주요 수법 항목들을 기준으로 문서간의 유사도를 측정하였다. 이외에도 [4]은 사모대출과 관련된 유사 판례를 검색하기 위해, 각 판결문마다 대출인과 차용인의 특성, 보증의 종류 등 사모 대출과 관련된 요소를 직접 추출하여 전체 문서를 대상으로 한 TF-IDF 값과 요소간 유사도를 통해 유사사건 데이터셋을 구축하였다.

본 논문에서는 판결문이라는 특수한 목적으로 작성된 텍스트를 기반으로 수사 전문가 입장에서 유사 사건을 도출하는 것을 목적으로, 수사 전문가들 유사성을 판단할 때 사용하는 기준을 사용해 유사 사건 도출 실험을 진행하였다.

살인죄는 강력 범죄 중에서 가장 정형화되어 있어 유사 사건 검색을 진행하기 비교적 명확하기 때문에, 살인죄를 대상으로 선정하였으며, ‘판결서 인터넷 열람 사이트’를 통해 2013년부터 2020년도까지 살인죄 판결문 4,042건을 수집하였다. 정규 표현식을 사용해, 피고인의 범행 사실을 명시하고 있는 ‘범죄사실’을 추출하여 유사 사건 검색에 활용하였다.

표 1 형사 판결문 대상 유사 사건 검색 실험 방안

| 실험 방안 | 임베딩 방안                    | 유사도 측정 방안              |
|-------|---------------------------|------------------------|
| 1     | Word Mover Distance (WMD) | WMD 자체 제공              |
| 2     | GloVe                     | Soft Cosine Similarity |
| 3     | KoBERT                    | Cosine Similarity      |

기준 사건을 선정하고 다양한 임베딩 방안을 활용해 4,042건 중 가장 유사한 사건 1,2위를 추출하여 모델간의 합의를 이끌어 내고자 하였다. 임베딩 방안으로는 단어 단위로 임베딩을 하는 Word Mover Distance (WMD)[5]와 Global Vector for Word Representation (GloVe)[6] 방식을, 문장 단위로 임베딩 하는 KoBERT 모델[7]을 활용하였고 유사도 측정 방식으로는 코사인 유사도와 소프트 코사인 유사도[8]를 사용하였다. WMD 모델은 해당 모델에서 자체적으로 유사도를 측정하는 방안을 제공하기 때문에 별도의 유사도 측정 방안을 사용하지 않았다. 단어 임베딩 방식과 문장 임베딩 방식의 모델들을 이용했음에도 불구하고, 모델마다 도출하는 유사 사건이 모두 다르다는 것을 확인할 수 있었다.

이후, 기준 사건과 각 모델의 1위와 2위 사건을 수사 관련 종사자 11명이 직접 순위를 매겨 유사 사건 결과를 평가하였다. 기준 사건에 따라 한 모델이 더 우수한 성과를 내는 경우는 있었으나, 전체 순위에서는 그 어떤 모델도 대다수의 동의를 이끌어 내지 못했다. 따라서, 단순히 임베딩 모델을 기반으로 유사 사건 도출하는 것은 형사 판결문에 적합하지 않다는 것을 파악하였다. 이에 전문가들의 유사 사건 평가 이유를 분석하여 그 선정 이유에 대해 파악하고자 하였다. 유사 사건 판단 이유로는 살해 방법, 동기, 흉기 등 범죄 수단과 관련된 단어가 많았으며, 가장 많이 언급된 분류가 유사 사건 판단에 가장 큰 영향력을 가지고 있다고 판단하여 빈도수로 정렬하였다.

표 2 유사 사건 판단 시 가장 많이 언급된 상위 3위 항목

| 분류           | 빈도수 |
|--------------|-----|
| 동기           | 35  |
| 범행 도구        | 21  |
| 피고인과 피해자의 관계 | 17  |

전체 113건의 결과 중에서 동기가 35건(31%), 범행 동기가 21건(18%), 피고인과 피해자의 관계는 17건(15%)이었다. 위 결과를 통해, 유사 사건을 평가할 때 범죄 수법을 주요 기준이 되며, 그 중 동기를 가장 많이 고려한다는 것을 알 수 있었다.

### 3. 동기 기반 분류 데이터셋 구축 및 실험 결과

#### 3.1 동기 기반 분류 데이터셋 구축

형사 판결문 범죄사실에는 피고인과 피해자의 관계와 흉기는 간단한 키워드로 서술되는 반면, 동기는 다양한 방식으로 서술되기 때문에 단순 키워드 매칭 검색이나 임베딩 값을 이용해 유사 사건 검색에는 어려움이 있다. 따라서 유사 사건 도출을 위해 범죄 사실 내 서술된 동기를 기준으로 판결문 분류 작업을 진행하였다. 살인죄에서 서술하고 있는 동기로 데이터셋을 구축하기 위해, 4,042건의 판결문 중 살인죄만 다루고 있는 판결문을 선별하였다. 범죄사실 내 동기를 나타내는 패턴을 수립해 정규 표현식을 이용해 동기를 나타내고 있는 문장인 동기 문장을 추출하였다. 동기 패턴이 존재하지 않는 판결문은 수작업으로 동기 문장을 추출하였다.

과학적범죄분석시스템 내 2015년부터 2019년까지 서울지방경찰청 관할 내에서 발생한 살인 및 치사 사건 총 216건의 데이터를 분석한 연구[9]에서는 동기를 경제적 이유, 분노/보복/우발 등, 치정, 정신병 등, 가정불화, 기타(불상)으로 분류하였다. 하지만, 실제로 살인죄 하급심 판결문을 분석한 결과 언어적인 세분화가 어려워 동기 분류를 치정, 원한, 정신, 금품으로 규정하여 앞서 추출한 동기 문장을 분류하였다. 예를 들어 치정으로 분류하는 조건은 1) 피고인과 피해자의 관계가 부부 또는 연인이며, 2) 피고인과 피해자의 관계로 인해 범행을 저지르는 경우이다. 피고인과 피해자 사이 채무 관계 또는 피고인이 금품을 노리고 범행을 저지른 경우, 금품으로 분류하였다. 피고인이 정신 질환 등의 이유로 범행을 저지른 경우, 정신으로 분류하였다. 이외 피고인과 피해자의 다툼으로 인해 범행이 일어난 경우 원한으로 분류하였다. 동기 문장 중 동기 분류를 잘 나타내고 있는 키워드를 추출해, 1,012건의 사건 내 동기 문장과 키워드로 구성된 데이터셋을 구축하였다.

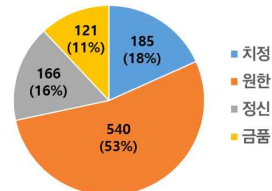


그림 1 동기 분류 데이터셋 내 동기 분포

### 3.2 동기 기반 분류 데이터셋 기반 실험 및 결과

동기 분류 데이터셋 내 키워드 데이터와 문장 데이터를 머신러닝 계열인 Decision Tree, Random Forest, SVM 모델에 학습시켰으며, 문장 데이터를 트랜스포머 계열인 KoBERT[7], KLUE(bert-base)[10] 모델을 학습하여 검증하였다. 문장 데이터를 처리하지 못하는 머신러닝 계열 모델에는 Konlpy Mecab 형태소 분석기[11]를 통해 명사만 추출하여 사용하였다. 각각의 동기 분류 데이터셋을 랜덤으로 8대 2 비율로 나눠 80% 데이터로 모델을 학습시키고, 20% 데이터로 모델의 성능을 검증하였다.

표 3 동기 분류 데이터셋 각 모델의 정확도

| 구분  | 모델               | 모델 종류       | 데이터셋 |
|-----|------------------|-------------|------|
| 키워드 | Decision Tree    | ML          | 0.74 |
| 키워드 | Random Forest    | ML          | 0.95 |
| 키워드 | SVM              | ML          | 0.91 |
| 문장  | Decision Tree    | ML          | 0.73 |
| 문장  | Random Forest    | ML          | 0.75 |
| 문장  | SVM              | ML          | 0.61 |
| 문장  | KoBERT           | Transformer | 0.82 |
| 문장  | KLUE (bert-base) | Transformer | 0.87 |

동기 분류 데이터셋에서 가장 좋은 성능을 보인 모델은 키워드 데이터를 활용한 Random Forest였다. 문장 데이터를 사용한 트랜스포머 계열 모델 중에는 KLUE(bert-base) 모델이 가장 좋은 성능을 보여주었다. 같은 트랜스포머 계열 모델임에도 불구하고 KLUE(bert-base) 모델이 KoBERT 모델 보다 좋은 성능을 보여준 이유는 KLUE(bert-base) 모델의 학습량이 KoBERT보다 많기 때문으로 추측된다(KLUE 모델의 사전 학습량: 문장 473M, 단어 615K, KoBERT 모델의 사전 학습량: 문장 5M, 단어 54M).

키워드 기반 Random Forest 모델이 좋은 성능을 보여준 것은 입력 값인 키워드가 동기 문장에 비해 많은 전처리 과정을 통해 동기 분류의 성격을 띠고 있는 키워드로만 구성되어 있기 때문으로 보인다. 또한, 사람에 의해 추출된 키워드 데이터 정도까지의 높은 정확도를 유지하기 위해서는 많은 데이터로 학습된 트랜스포머 계열의 모델을 사용해야 하는 것으로 보인다.

### 4. 결론

본 논문에서는 형사 하급심 판결문을 수집해 유사 사건 검색을 진행하였으나, 모델 간의 합의를 이끌어 내기 어려워 사람을 통한 평가를 시도하였다. 그러나 사람 간의 합의 또한 명확하지 않아, 유사 사건 판단 기준을 분석하였다. 평가자들은 흥기나 살해 방법 등을 주요 정보로 판단하였으나 그보다 동기를 더 중요한 기준으로 인식하고 있었다. 판결문 범죄사실 내 서술되어 있는 동기를 확인하였고, 여러 방식으로 서술되어 있는 동기를 치정, 원한, 금품, 정신으로 분류하였다. 범죄사실 내 동기를 서술하고 있는 문장과 해당 문장 내 키워드를 추출하여 데이터셋을 구축하였다. 동기 분류 데이터셋은 다양한 머신러닝 모델과 트랜스포머 계열 모델을 통해 검증한 결과, 문장 데이터는 KLUE(bert-base) 모델이, 키워드 데

이터는 Random Forest 모델이 가장 좋은 성능을 보여주었다. 하지만, 키워드 데이터의 경우, 사람에 의해 추출된 데이터므로 사람이 직접 추출한 키워드만큼의 성능을 도출하고자 한다면 많은 양의 데이터로 학습된 트랜스포머 계열의 모델을 쓰는 것이 적합한 것을 확인할 수 있었다. 추후, 동기뿐만 아니라 다양한 범죄 수법을 고려하여 유사 사건 검색을 유의미하게 진행할 수 있는 연구를 진행할 것이다.

### 참고 문헌

[1] 홍수정, “[법의날 특집] 주목받는 리걸테크 산업”, 2021년 4월 26일. <https://www.lawtimes.co.kr/Legal-News/Legal-News-View?serial=169586> (접근된 2023년 4월 11일).

[2] 강예지, 강혜린 and 박서운 and 장연지 and 김한샘. “딥러닝 알고리즘을 이용한 유사 판례 매칭 데이터셋 구축 방안 연구”, 한글 및 한국어 정보처리 학술대회 논문집, pp.72-76, 2021

[3] 강명중, 강일구 and 김민정 and 조성호 and 이경은 and 이미희. “침입절도 범죄 간 유사도 측정 알고리즘의 개발 및 응용에 관한 연구.” 한국경찰학회보,20(6),57-82, 2018.

[4] Xiao, C., Zhong, H., Guo, Z., Tu, C., Liu, Z., Sun, M., ... & Xu, J. (2019). Cail2019-scm: A dataset of similar case matching in legal domain. arXiv preprint arXiv:1911.08962.

[5] M. J. Kusner, Y. Sun, N. I. Kolkin와/과K. Q. Weinberger, “From Word Embeddings To Document Distances”.

[6] J. Pennington, R. Socher와/과C. Manning, “Glove: Global Vectors for Word Representation”, in Proceedings of the 2014 Conference on EMNLP, 2014, pp 1532-1543. doi: 10.3115/v1/D14-1162.

[7] “KoBERT”. <https://github.com/SKTBrain/KoBERT>.

[8] G. Sidorov, A. Gelbukh, H. Gomez-Adorno와/과D. Pinto, “Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model”, Comput. Syst., vol 18, 호 3, 9 2014, doi: 10.13053/cys-18-3-2043.

[9] 박정준, 이수정 and 홍정윤. “서울지역 살인 혐의(살인 및 치사) 사건의 범행 수법에 따른 유형화 연구 - SCAS 자료에 대한 잠재계층분석의 적용 -” 한국경찰연구 19, no.3 (2020) : 87-112.

[10] S. Park기타, “KLUE: Korean Language Understanding Evaluation”. arXiv, 2021년 11월 2일. 접근된: 2023년 4월 11일. [Online]. Available at: <http://arxiv.org/abs/2105.09680>

[11] 박은정와/과조성준, “KoNLPy?: 쉽고 간결한 한국어 정보처리 파이썬 패키지”. 제 26회 한글 및 한국어 정보처리 학술대회 논문집, 2014년.