

뉴스 기반 유사 판례 검색 및 형량 분석 서비스

이정원⁰, 박상근

경희대학교 소프트웨어융합학과

spwldmlgla@naver.com, sk.park@khu.ac.kr

News-Based Similar Case Search and Sentence Analysis Services

Jeongwon Lee⁰ Sangkeun Park

Department of Software Convergence, Kyung Hee University

요 약

온라인에서 특정 범죄에 대한 구형 및 판결에 대한 뉴스를 자주 접하게 된다. 일반인은 뉴스에서 다루는 범죄와 해당 범죄에 대한 구형 및 판결이 적절한지 알기 어렵다. 본 연구는 일반인도 특정 범죄에 대한 구형 및 판결이 적절한지 알기 쉽도록 뉴스 기사 텍스트를 기반으로 유사 판례를 검색하고, 그 결과를 시각화하여 사용자의 이해를 돕는 크롬 플러그인 서비스를 개발했다. 이 서비스는 자동화된 유사 판례 검색 기능과 형량 분포 시각화, 빈도 기반 통계적 해석을 제공하여 데이터에 기반한 객관적인 형량 평가를 가능하게 한다. 이를 통해 사용자는 특정 조건의 범죄에 따른 판결 분포를 쉽게 이해하고 법적 일관성을 쉽게 확인할 수 있다.

1. 서 론

온라인 뉴스를 통해 특정 사건에 대한 법원의 판결 결과를 접하는 대중들은 해당 판결이 객관적이고 공정한 판결인지 의문을 갖는 경우가 많다[1]. 이러한 실제 판결과 국민 법감정과의 괴리감이 발생하면서 대중들은 사건의 판결 근거를 이해하는데 어려움을 겪고 있으며, 이는 나아가 사법 불신으로까지 이어질 우려가 있다.

특정 사건이 발생하면 대중들은 해당 사건에 어느 정도의 형량이 나올지 궁금해하며, 판결이 내려지기 전부터 다양한 예상과 논의를 벌인다. 이 과정에서 사람들은 해당 사건에 어느 정도의 형량을 예상하기 위해 양형 기준을 찾아보기도 하고, 기존 유사 사건에 대한 판례를 참고하기도 한다. 하지만 법률 용어에 익숙하지 않은 일반인은 특정 사건과 유사한 판례를 검색하는데 어려움을 겪을 수 있다.

일반인이 판례를 쉽게 검색할 수 있도록 하기 위한 다양한 연구가 수행되었다. 법률 용어를 잘 모르더라도, 사용자가 입력한 단어와 유사한 단어까지 고려해서 판례를 찾아주거나[2, 3], 판례 검색 결과를 사용자가 이해하기 쉽게 핵심 단어 및 문장을 강조해 주는 연구[2, 5, 6]가 수행되었다. 하지만 이러한 연구는 사용자가 키워드를 기반으로 판례를 검색하고, 해당 검색 결과가 왜 나왔는지 강조하는데 초점을 맞춰기 때문에, 판례를 기반으로 사용자가 원하는 특정 사건에 대한 형량 분포 등을 한눈에 파악하기는 어렵다는 한계가 있다.

이에 본 연구는 키워드, 판결문 검색 기반이 아닌 뉴스 기사를 기반으로 유사 판례를 검색하고, 해당 문서에서 다루는 사건과 유사한 사건의 판례들을 사용자가 이해하기 쉽게 분석 및 시각화하여 제공하는 크롬 플러그인 서비스를 제안한다.

2. 관련 연구

2.1. 법률 검색 방법

법률 용어를 잘 모르는 일반 사용자도 어려움 없이 판례 관련 정보를 검색할 수 있도록 돕기 위한 다양한 연구가 수행되었다. 법률 용어를 잘 모르는 사람이 법률 용어가 아닌 일반 키워드로 검색을 하더라도 관련된 판례가 잘 검색되도록 하기 위한 단어 임베딩 기법을 제안한 연구가 있었으며 [2, 3], 특정 판례를 입력하면 관련 법령을 빠르게 찾아주는 Law2Vec을 알고리즘을 제안한 연구도 있었다 [4]. 기존 연구는 사용자가 특정 판례 관련 정보를 손쉽게 도와주지만, 이를 위해서는 본인이 직접 키워드나 관련 판례를 찾아 입력해야 하는 한계가 있다.

2.1. 법률 검색 결과 제공

일반 사용자는 법률 용어에 익숙지 않으므로, 사용자가 이해하기 쉬운 형태로 검색 결과를 제공하는 것이 중요하다. 일반적으로 빅케이스¹, 로앤비² 등의 법률 검색

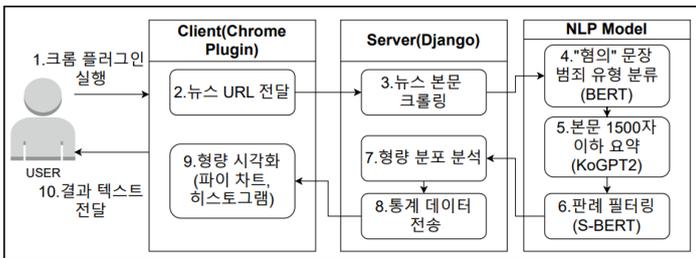
¹ <https://bigcase.ai/>

² <https://www.lawnb.com/>

서비스에서는 판례 검색 결과가 텍스트 제공된다. 사용자의 검색 결과 이해를 돕기 위해, 내가 검색한 키워드와 유사 단어까지 강조해서 텍스트 기반의 판례 검색 결과를 제공하는 연구가 수행되었다. 사용자가 입력한 키워드와 일치하거나 유사한 단어를 색상으로 강조[2, 5]하거나 굵은 글씨로 강조[6]한 연구가 있었다. 하지만 이런 텍스트 기반의 검색 결과 제공 방식은 사용자가 특정 사건에 대한 형량 분포를 한눈에 파악할 수 없다는 한계가 존재한다. 사용자가 찾는 판례들을 기반으로 형량 분포를 시각화해서 보여준다면 사용자가 검색한 사건에 대해 과거의 판결을 한눈에 파악할 수 있다.

본 연구에서는 기존 연구의 한계를 극복하기 위해, 키워드 기반의 검색이 아닌 온라인 뉴스 문서 기반의 판례 검색을 지원하고, 검색 결과로 나온 판례들을 종합하여 전체적인 형량 분포를 분석 및 가시화된 결과를 보여주는 서비스를 개발했다. 이를 통해, 사용자가 특정 범죄에 대한 과거의 판결을 쉽게 이해할 수 있고, 나아가 유사 범죄에 대해 어떤 판결이 내려질지 미리 예측해 볼 수 있다.

3. 사건 뉴스 형량 분석 서비스 개발



[그림 2] 사건 뉴스 형량 분석 시스템 구조

전체 서비스 구조는 [그림 2]와 같다. 사용자가 범죄 형의가 포함된 네이버 뉴스 기사를 크롬 플러그인의 “뉴스 형량 분석” 버튼을 클릭해서 해당 뉴스에 대한 분석을 요청하면, 해당 페이지의 URL이 서버로 전송된다 [그림 3].



[그림 3] 해당 기사 관련 판례 분석 요청

Django REST Framework로 구축된 서버는 전달받은 URL에서 뉴스 본문의 텍스트만 크롤링한다. 해당 본문 텍스트를 기반으로 범죄 유형을 판단 및 해당 범죄 유형의 유사 판례를 찾고, 판결 분포를 시각화한다.³

³ <https://youtu.be/USI90FsGTmc>

3-1. 판례 데이터 수집 및 전처리

Hugging Face에 공개된 85,659개의 판결문 데이터⁴를 수집했다. 해당 데이터는 각 판례에 대한 범죄 유형 데이터가 없기 때문에 검찰청 범죄통계표⁵를 기준으로, 해당 판결문 데이터의 판례의 사건명과 범죄 유형별 핵심 키워드를 매칭해서 78개의 범죄 유형 정보를 추가했다. 판결문의 형량 정보를 정량적으로 분석하기 위해, 정규 표현식을 사용하여 판결문의 징역 및 집행유예 정보는 년/월 단위의 숫자로 추출하고 벌금은 원 단위로 추출했다.

3-2. 범죄 유형 예측 및 유사 판례 필터링

사전에 구현한 범죄 유형 예측 BERT 모델을 활용해서 사용자가 전송한 뉴스 기사가 어떤 범죄 유형을 다루고 있는지 분류한다. 범죄 유형은 검찰청 범죄통계표⁵에서 분류하고 있는 78개 범죄 유형을 활용했다. 예를 들어, 뉴스 기사에서 폭행, 상해, 협박, 공갈 등이 언급되고 있다면 해당 기사는 “폭력” 유형으로 분류되며, 부당이득, 보험사기 등이 언급되고 있다면 해당 기사는 “사기” 유형으로 분류된다. 85,659개의 판결문 데이터 중에서, 해당 범죄 유형에 해당하는 판례만 필터링하고, 이 중에서 뉴스 본문 텍스트와 유사한 판례를 찾는다. 이 과정을 위해서, 한국어 GPT2 모델⁶을 활용해 뉴스 기사 본문 텍스트를 1500자 이내의 요약문을 생성한다. 그 다음, S-BERT 기반으로 사전 구축한 임베딩을 활용해서 뉴스 기사의 요약문과 코사인 유사도가 0.7 이상인 판례만 필터링한다.

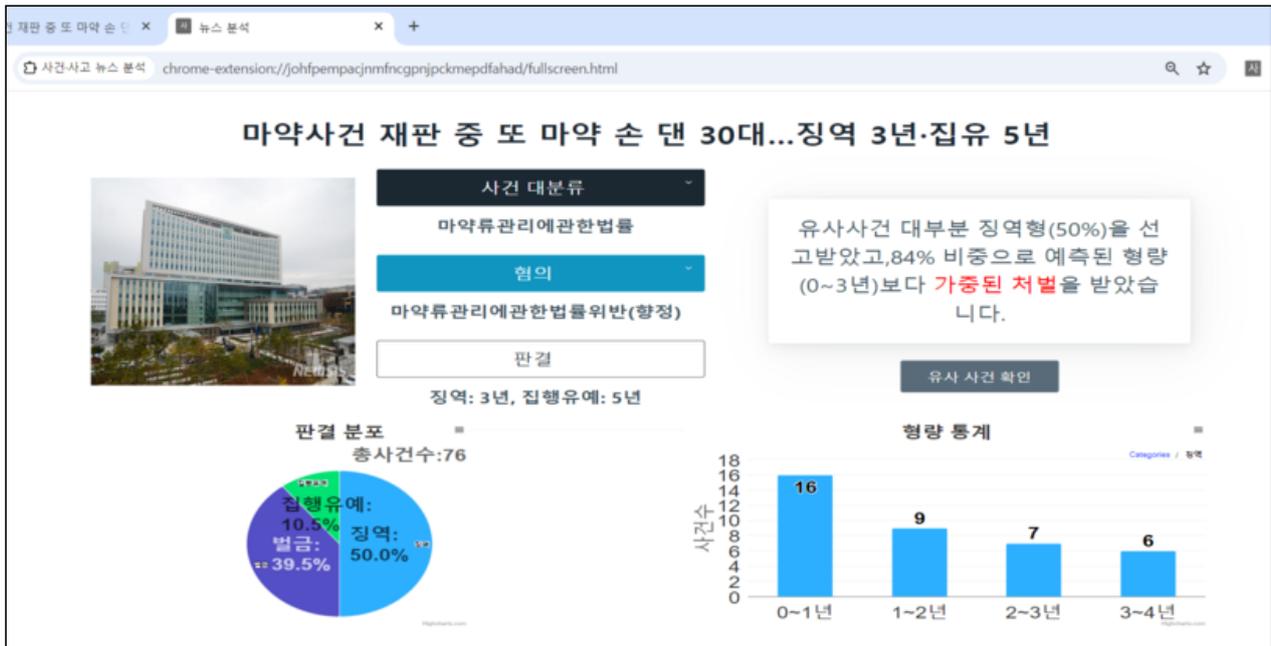
최종적으로 사용자가 요청한 뉴스 기사와 가장 유사하다고 판단된 판결문을 분석하고 시각화한다. [그림 4]. 유사 판결문들의 형량을 사용자가 이해하기 쉽게 시각화하기 위해 Highchart.js 라이브러리⁷를 활용했다. 형벌 종류(징역, 집행유예, 벌금)의 분포를 파이 차트로 표현해 전체적인 분포를 파악할 수 있다. 사용자가 형벌 종류 중 하나를 클릭하면, 해당 형벌에 대한 세부 형량을 히스토그램으로 시각화해서 그 분포를 확인할 수 있다. 형량은 징역, 집행유예, 벌금으로 분류된다. 징역과 집행 유예는 연 단위로 분류하고, 벌금은 100만원 단위로 분류해서 히스토그램으로 나타냈다.

⁴ https://huggingface.co/datasets/joonhok-exo-ai/korean_law_open_data_precedents

⁵ <https://www.spo.go.kr/site/spo/crimeAnalysis.do>

⁶ <https://huggingface.co/skt/kogpt2-base-v2>

⁷ <https://www.highcharts.com>



[그림 4] 사용자가 요청한 뉴스에서 다루는 범죄에 대한 판결 분석 결과

이렇게 유사 판결문을 기반으로 형량의 분포를 확인하면 특정 사건에 대한 판결의 일관성을 간접적으로 확인해 볼 수 있다. 전체 형량 분포를 설명하는 텍스트도 함께 제공해서 사용자의 이해도를 높였다.

4.결론 및 제언

본 연구에서는 사용자가 법적 전문 지식 없이도 온라인 사건 뉴스의 유사 판례를 쉽게 검색하고, 그에 대한 형량 분포를 쉽게 이해할 수 있는 서비스를 개발했다. 이를 통해 사용자는 특정 사건에 대해 대체로 어떤 형량이 부과되었는지 파악하고, 해당 사건에 대한 구형 및 판결이 일반적인지 특수한 경우인지 확인할 수 있다. 향후 연구로, 사건의 배경, 가해자의 인적사항 등 다양한 변수를 고려한 분석을 수행해서 특정 사건에 대한 구형 및 판결을 사용자가 더 정확하게 이해하는데 도움을 줄 수 있다.

참고 문헌

[1] 박미영, 고무줄 형량 줄었지만... ‘국민 법감정과 괴리’ 해소는 숙제 [심층기획 - 출범 15주년 맞은 대법 양형위]. 세계일보, <https://www.segye.com/newsView/20220425516011>, 2022.

[2] 박길식, "S-BERT와 유사도 기반 파인튜닝을 적용한 개선된 판례 검색", 2023.

[3] 박구락 et al. "조세심판 문서 검색 효율 향상 모델에 관한 연구", 한국융합학회논문지, 제10권, 제6호, pp. 41-47, 2019.

[4] 김나리 & 김형중, "연관법령 검색을 위한 워드 임베딩 기반 Law2Vec 모형 연구" 디지털콘텐츠학회논문지 18.7 pp.1419-1425 (2017) : 1419.

[5] 조희진 & 강윤철, "법을 인공지능 연구 활성화를 위한 판례 데이터셋 자동 라벨링 및 BERT를 활용한 판례 유사성 분석", 아시아태평양융합연구교류논문지, 제9권, 제10호, pp. 95-109, 2023.

[6] 강예지 & 김한샘, "한국어 법률 텍스트 처리 방안 연구 - 유사 판례 도출 태스크를 중심으로 -", 언어사실과 관점, 제58권, pp. 453-473, 2023.