

정보통신관련법 및 판례 관련 지원 서비스

Information and Communication Act and the related precedents support services

요 약

법조계 종사자 및 정보통신분야 종사자 간의 간극 해소를 위하여 법, 판례 및 정보통신용어 연계 서비스를 제안한다. 링크 예측 및 키워드 추출 기법을 활용하여 법 및 판례와 정보통신용어 간의 연결관계를 밝히고 이를 기반으로 직관성 높은 서비스를 제공한다.

1. 서 론

2020년 12월 정식 서비스를 시작하였던 스캐터랩(SCATTER LAB)의 열린 주제 대화형 인공지능 챗봇(Open-Domain Conversational AI Chatbot)인 이루다는 여러 사회적 파장을 일으켰지만, 그 중 직접적으로 법적인 제재를 받았던 이유는 개인정보보호법 제28조의2제2항을 위시한 8개 조항을 위반하였기 때문이다[1]. 비단 이루다의 경우뿐만 아니라, 현재 제4차 산업혁명의 최전선을 담당하는 인공지능을 연구하는 데에 있어 연구 데이터는 정보통신정책과 법에서 자유로울 수 없다. 위반했을 경우의 처벌 수위 역시 마찬가지인데, 현행 징벌적 손해배상의 범위인 3배 배상제도의 수위를 사건의 경중에 따라 강화해야 한다는 목소리도 있으며[2], 한편으론 강력한 제재는 산업 전반의 발전 저해를 초래한다는 목소리가 동시에 있어 양측의 의견을 적절히 절충할 필요성이 있다.

법은 필연적으로 기술 발전 속도에 발빠르게 대처하지 못한다. 산업 현장에서 특정 기술이 빠르게 등장하고 사라지는 탓도 있겠지만, 그 말은 그만큼 기술이 발빠르게 발전한다는 뜻이다. 기술 발전 속도에만 맞춰서 법을 제정하자니 산업 현장의 목소리를 제대로 반영할 수 없고, 그렇다고 모든 이해관계를 고려한 법을 도입하자니 역으로 산업 현장과 그 기술로 인하여 피해를 본 사람들을 구제할 수 없는 문제가 발생한다. 하지만, 법을 이해하고 전공하는 사람과 산업을 이해하는 사람은 그 이해의 깊이가 다를 수밖에 없으나, 가장 기초적인 단계에서부터 정보 및 의견의 교류가 가능한 체계가 존재한다면 이 문제는 이전보다는 발빠르게 대처할 가능성이 높아진다. 따라서 산업 현장, 특히 정보통신 분야 종사자 및 법조계의 이해 간극을 좁히기 위하여 별도로 존재하던 법 및 정보통신 관련 정보를 하나로 모을 수 있는 플랫폼을 제안한다.

2. 관련 연구

현재 대한민국 법원의 종합법률정보 포털을 제외하고 양질의 판례를 제공하는 대중성 있는 플랫폼은 약 4개가 있다[3]. 저작권법 제7조에 의거하여, 법률 및 판례는 저작권법의 보호를 받지 못한다. 하지만 프로젝트 진행에 있어서의 문제를 최대한 피하기 위해, 판례 수집은 대한민국 법원 종합법률정보에서 조회 가능한 정보통신 분야의 판례에 한한다. 해당되는 판례 및 관련법 그리고 정보통신기술 및 그 용어를 수집, 데이터베이스를 구축하여 각 노드 간의 관계를 가장 잘 나타낼 수 있는 그래프 데이터베이스 및 링크 예측 기술을 활용하여 현재 존재하는 정보 간의 관계 및 추후 등장할 기술의 예상 논점을 쉽게 파악할 수 있도록 한다.

가능한 키워드는 수집된 판례에서 최대한 추출한다. 추출된 결과 중 정보통신에 관련된 키워드는 정보통신용어 DB의 유관 키워드 및 인접 노드 간의 유사도 비교를 통하여 최대한 비슷한 경우를 추천할 수 있도록 한다. 추가로 판례에 등장하지 않거나, 혹은 잠재적 법적 분쟁 가능성이 있는 기술에 대한 저촉 가능성을 예측하며, 더 나아가 특정 기술을 기준으로 관련법 및 판례를 제시할 수 있도록 구현한다.

해당 플랫폼을 웹을 통해 구현한다면, 조회수를 별도로 계산하여 해당 시점에서 가장 많은 조회수를 기록한 용어나 법, 판례를 제시하여 가장 관심도가 높은 부분에 대해 파악할 수 있도록 구현한다. 가능하다면 댓글 기능을 추가 구현하여 해당 기술 및 법에 대한 자유로운 의견 교류를 장려할 예정이나, 프로젝트를 진행함에 있어 불필요한 관리 소요가 발생할 가능성이 커 우선도를 낮게 배정한다.

2.1. 링크 예측

링크 예측(Link Prediction)이란 현재시점의 네트워크가

주어졌을 때, 미래시점의 네트워크에 추가될 엣지를 예측하는 문제를 의미한다. 또한 현재 관계가 나타나지 않은 노드 간의 연결성을 예측하는 문제 역시 포함한다[4]. 노드를 단어의 개념으로 인식한다면 자연어 처리 분야의 Word2Vec 기술과 유사하다고 볼 수 있으나, 선형적인 형태의 자연어와는 별개로 그래프는 매우 다양한 방향성을 띌 수 있다.

예측 방법에는 Translation-based model 및 Neural-based model이 존재한다. Translation-based model의 경우 Entity와 Relation을 모두 벡터로 표현하며, Relation은 Entity의 위치를 다른 위치로 옮겨주는 연산자 역할로 간주한다. 따라서 주체 Entity(h)와 객체 Entity(t) 그리고 이들 사이의 관계를 나타내는 Relation(r)으로 이루어진 트리플 <h, r, t>로 두었을 때 $h + r = t$ 와 같이 귀결되는 경우로써 단순하지만 Neural-based보다 우수한 성능을 보인다. 하지만 1대1 대응이 아니면 표현할 수 없다. Neural-based model의 경우 하나의 트리플 <lhs, rel, rhs>에 대해 2개의 임베딩 표현을 형성한다. 그리고 그 두 표현 사이의 유사도를 이용하여 스코어 함수를 정의한다. 낮은 계산 복잡도를 지녀 대규모 지식 그래프에 적용 가능하다.

2.2. 판례 제공 서비스

현재 판례 제공 플랫폼의 서비스 양상은 전체 공개되지 않은 판결문에 대한 서비스가 주가 된다. 독보적인 판례를 확보하여 타 플랫폼과의 차별점을 강조하는 서비스가 대부분이며, 빅케이스의 경우 단순 검색이 아닌 쟁점별 판례 검색이나 핵심요약, 유사도 높은 판례를 제공하는 등 NLP를 도입한 검색 서비스가 특징이다. 다만, 아직 특정 산업군에 특화된 판례 서비스는 전무하다. 따라서, 특정 산업군을 겨냥한 서비스는 충분한 가치가 있을 것이다

3. 문제 정의

현재, 판례의 사건명은 판결사항, 판결요지, 판례내용 등을 담당자가 확인하여 직접 명명하는 방식으로 이루어진다. 따라서, 명명하는 자에 따라 저촉되는 다양한 법이나 사건의 요지가 드러나기도, 혹은 간략한 쟁점이나 핵심만이 나타난다. 따라서, 본 프로젝트에서 비 법조계 인원이 관련 정보통신 기술명이나 법령 조항만으로 손쉽게 관련 판례를 찾아내기 위해서 판결요지 및 판례내용으로부터의 핵심 키워드 추출이 필요하다.

또한, 핵심 키워드를 통해 정보통신 유관 판례로 분류된 판례를 직접 명시된 법과 정보통신기술과 연결하고, 추가로 링크 예측을 통해 유관 판례를 제시한다. 이 때, 직관성을 위해 가장 인접도가 높은 10개의 노드를 추출하고, 사용자 편의에 따라 최소 5개에서 최대 10개까지 표시할 수 있도록 한다.

법령 조항 및 정보통신용어를 사용자에게 제시할 때엔 해당 노드에 관련있는 판례가 몇 건인지를 표시할 수 있도록 하며, 판례를 사용자가 조회할 땐 해당 판례와 관계있는 법령과 정보통신용어의 목록을 확인할 수 있도록

한다. 판례번호를 알고 있을 경우 더 쉬운 방향으로 조회를 할 수 있도록, 판례를 조회할 경우의 URL 파라미터에는 판례번호를 포함할 수 있도록 하며, 판례번호로 검색 역시 가능하도록 한다. 검색 속도 개선을 위한 알고리즘은 본 연구에서 고려하지 않는다.

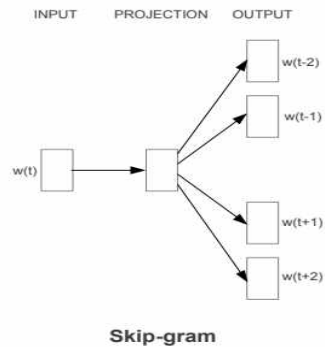
4. 정보통신관련법 및 판례 관련 지원 서비스 설계

4.1. 판례 분석

본 연구에선 법제처에 등록된 판례 및 정보통신관련법을 기반으로 데이터베이스를 구축한다. 2022년 10월 기준 84,980건의 판례가 존재한다. 하지만, 판례명을 기준으로 정보통신분야와 충분히 연관성이 있다고 볼 수 있는 '정보통신', '개인정보', '저작권', '컴퓨터', '인터넷'과 같은 용어를 색인 기준으로 선정하여 분류를 진행하였을 때 반환된 판례는 총 630건으로, 총 판례의 1%가 채 되지 않는다. 따라서, KeyBERT 및 TF-IDF를 통한 핵심 키워드를 추출한다. 그 후 추출된 키워드를 기반으로 핵심 정보통신용어와의 단어 유사도 비교 및 수동 분류를 통해 정보통신 연관 판례임을 판별한다.

4.2. 단어 유사도 분석 모델

Word2Vec은 단어를 임베딩하여 벡터 유사도 검증을 가능토록 하는 모델이다. 따라서, 우리말샘 사전의 학습 데이터로 Word2Vec의 Skip-gram 모델을 학습하여 생성한다. Skip-gram 모델은 하나의 입력값을 기반으로 입력값과 유사한 다수의 출력값을 예측하는 모델이다.



[그림 1] Skip-gram 구조

학습 데이터는 한국정보통신기술협회의 정보통신용어 사전을 토대로 구축한다. 모델 구축 전에, 확보한 정보통신용어를 형태소 분류기에 등록하여 정상적인 단어가 형태소 단위로 쪼개져 발생하는 문제점을 방지하도록 한다. 학습을 토대로 완성된 모델을 기반으로 판례에서 추출한 키워드와의 의미 유사도 비교를 진행한다.

5. 결론 및 향후 연구

정보통신 분야 종사자는 구현하고자 하는 프로젝트나 연구가 법적으로 어떠한 문제가 발생할 수 있는지에 대해서 더 쉽게 알 수 있고, 법 관련 종사자 역시 정보통

신 분야의 지식 및 쟁점을 자신들의 연구에 더 잘 활용할 수 있을 것이라 기대한다. 또한, 구축한 링크 예측 결과 및 그래프를 토대로, 새로운 판례가 추가되거나 신기술의 등장, 법의 개정 등 상황이나 맥락이 달라질 경우에도 더 유연하게 정보를 제공할 수 있을 것으로 전망한다. 추가로, 기술 교육을 받는 인원이나 해당 기술을 독학으로 접하는 사람들이 일으킬 수 있는 개인정보 침해나 기타 연구윤리에 반하는 행위를 예방할 수 있다고 전망한다.

향후 연구로는, 정보통신 관련 기사나 커뮤니티의 쟁점을 추가하여 더 복합적인 정보 취득이 가능한 방향으로 진행할 수 있다. 또한, 해당 데이터를 기반으로 교육 콘텐츠를 생산하거나, 더 나아가 산업 현장을 고려한 관련법 제정을 기대할 수 있다.

참고문헌

- [1] 「개인정보위, ‘이루다’ 개발사 (주)스캐터랩에 과징금·과태료 등 제재 처분」, 『개인정보보호위원회』, 2021년 4월 28일.
- [2] 「개인정보 유출로 인한 손해배상책임의 최근 동향」, 『보안뉴스』, 2016년 7월 7일.
- [3] 「온라인 기업 ‘판례 검색 사이트’ 본격 경쟁 돌입」, 『법률신문』, 2022년 5월 12일.
- [4] David Liben-Nowell · Jon Kleinberg, 「The Link Prediction Problem for Social Networks」, 『CIKM』 제 12회, Association for Computing Machinery, 2003.
- [5] Sanket Doshi, 「Skip-Gram: NLP context words prediction algorithm」, 『Towards Data Science』, 2019년 5월 17일.