

Doc2Vec 문서 임베딩을 이용한 질의문과 판례 자동 연결 방안 연구

강예지[○], 강혜린, 박서윤, 장연지, 김한샘[†]
연세대학교 언어정보학협동과정, 언어정보연구원[†]

{yjkang5009, hyerink, seoyoon.park, yeonji3547, khss[†]}@yonsei.ac.kr

A Study on the Connecting Method of Query and Legal Cases Using Doc2Vec Document Embedding

Ye-Jee Kang[○], Hye-Rin Kang, Seo-Yoon Park, Yeon-Ji Jang, Han-Saem Kim[†]
Interdisciplinary Graduate Program of Linguistics and Informatics, Yonsei University

요 약

법을 전문 지식이 없는 사람들이 법률 정보 검색을 성공적으로 하기 위해서는 일반 용어를 검색하더라도 전문 용어가 사용된 법령정보가 검색되어야 한다. 하지만 현 판례 검색 시스템은 사용자 선호도 검색이 불가능하며, 일반 용어를 사용하여 검색하면 사용자가 원하는 전문 자료를 도출하는 데 어려움이 있다. 이에 본 논문에서는 일반용어가 사용된 질의문과 전문용어가 사용된 판례를 자동으로 연결해 주고자 하였다. 질의문과 연관된 판례를 자동으로 연결해 주기 위해 전문용어가 사용된 전문가 답변을 바탕으로 문서 분류에 높은 성능을 보이는 Doc2Vec을 이용한다. Doc2Vec 문서 임베딩 기법을 이용하여 전문용어가 사용된 전문가 답변과 유사한 답변을 제안하여 비슷한 주제의 답변들끼리 분류하였다. 또한 전문가 답변과 유사도가 높은 판례를 제안하여 질의문에 해당하는 판례를 자동으로 연결하였다.

주제어: Doc2Vec, 문서 임베딩, 문서 분류

1. 서론

온라인상에 다양한 전문 분야와 관련된 데이터가 공개됨에 따라 전문 지식이 없는 일반인이 전문 자료를 쉽게 검색하고 활용할 수 있게 되었다. 그중 특히 법률적 도움이 필요한 사람들이 온라인상에 법률 전문 자료 검색이 가능해짐에 따라 변호사 도움 없이 스스로 소송을 준비하는 경우가 늘고 있다. 한국 경제[1]에서는 건수로 따지면 매년 60만~80만 건의 소송이 변호사 선임 없이 진행되었으며, 특히 소송 가액 300만원 이하의 소액 사건에서 ‘나홀로 소송’이 두드러졌다고 하였다. 실제 대부분의 사람들이 민사 소액 사건의 경우, 변호사의 도움 없이 포털 사이트에 질의하거나 본인의 케이스와 비슷한 케이스의 판례를 검색하여 기존 재판 사례를 참고하기도 한다.

하지만 판례는 일반인이 일상에서는 사용하지 않는 법률 전문용어로 구성되어 있어, 법률 지식이 없는 사람들이 사용하는 일반 용어로는 법령 정보를 검색하는 데 어려움이 있다. 또한 현 대법원 판례 검색 시스템은 해당 키워드가 포함되어 있고 판례 본문에 키워드가 자주 등

장하는 판례를 최상위 검색 결과로 제시하고 있는데, [2]에서 언급한 바와 같이 이러한 키워드 기반 검색 시스템은 사용자 선호도 검색이 불가능하여 사용자 검색 의도에 부합하는 정확한 검색 결과를 제공할 수 없다. 따라서 일반인이 온라인상에 공개되어 있는 법률 정보를 성공적으로 검색하기 위해서는 사용자가 일반 용어를 사용해도 법률 정보가 검색될 수 있어야 하며, 사용자 검색 의도에 부합하는 맞춤형 판례가 도출될 수 있도록 해야 한다. 이에 본 논문에서는 법률 전문 지식이 없는 사용자가 일반 용어를 사용한 질의문에 대해 전문가 답변을 매개로 질의문과 연관 관계에 있는 판례를 자동으로 도출해 낼 수 있도록 하기 위한 방법을 제시하고자 한다.

그간의 연구에서 Doc2Vec은 문서 분류, 감정 분석 등에서 높은 성능을 보여 왔다[3]. 본 연구에서는 Doc2Vec 문서 임베딩 기법을 이용하여 답변 데이터와 판례 데이터를 임베딩하고, 문서 간 유사도를 측정하여 질의문과 판례의 자동 연계 실험을 진행하고 그 결과를 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 법률정보 검색을 위한 판례를 이용한 연구와 논문에서 사용된 Doc2Vec 문서 임베딩을 활용한 관련 연구를 소개한다. 3

장에서는 본 논문에서 제안하는 질의문과 판례 데이터를 연결하는 모델을 설명하고, 4장에서는 모델을 기반으로 실험을 진행하였다. 5장에서는 결론과 향후 연구 방향에 대해 논의한다.

2. 관련 연구

2. 1. 법률 정보 검색

일반용어를 기반으로 한 법률 정보 검색 관련 연구는 일반용어에 대응하는 법률용어를 찾는 과정을 법률 온톨로지를 구축하는 방향으로 진행하거나[4], [5]에서처럼 웹 마이닝 기법을 이용하여 질의문-법령 매핑(mapping) 확률 생성 방법을 제안하며 ‘법령정보검색 시스템’을 제시하였다. 또한 [6]에서는 질의문과 판례를 연결하기 위해 Clustering과 Topic Modeling 기법을 사용하여 문서의 토픽을 정의하고, 군집화한 문서들을 주제별로 연결시켰다. 그런데 이전 연구의 법률 온톨로지에 포함된 법률용어는 법률정보 검색 시스템에서 인식이 불가능한 일반용어의 대체 수준으로, 사전에서 법률용어로 분류되어 있는 실제 법률용어가 아니다. 그리고 [6]의 연구에서는 토픽 내 등장하는 단어를 기반으로 연구자가 직접 토픽을 정의하는 과정에서 연구의 객관성이 떨어지는 한계가 있었다. 따라서 토픽 선정의 객관성을 보장 받으면서 보다 전문적인 결과를 얻기 위해서는 문장 수준에서의 임베딩 기법이 필요하다.

2. 2. Doc2Vec 문서 임베딩

단어나 문서가 얼마나 유사한지 계산하기 위해서는 이를 벡터화 해야 하는데, 이렇게 특정 언어 단위를 벡터로 표현하는 것을 임베딩이라고 한다. 비슷한 문맥을 가진 단어는 비슷한 의미를 갖는 경향이 있다는 [7]과 [8]의 분포 가설에 따라 유사도를 다루는 대표적인 방법론 중 단어 층위에서 Wodr2Vec 모델[9]이 distributed word representation 방법으로 비슷한 문맥을 지니는 단어를 비슷한 벡터로 표현한다. 이처럼 벡터로 표현된 단어들은 단어 사이의 거리를 계산할 수 있으며, 단어 간 거리를 통해 단어의 의미를 추론할 수 있다. Doc2Vec은 Word2Vec의 방식을 확장하여 문서 층위에서 문장, 단락, 문서 등을 벡터로 표현하며, 이를 통해 문서 간 유사도를 계산할 수 있도록 하여 문서 분류의 성능을 높인다[10]. Doc2Vec의 훈련 방식은 Word2Vec의 CBOW와 Skip-gram 훈련 방식과 유사한 PV-DM과 PVDDBOW가 있다. PV-DM은 문서 벡터와 단어 벡터를 모두 이용하여 다음에 나올 단어를 예측하는 것으로, 정해진 사이즈의 단어들을 문맥정보로 사용한다. 반면, PVDDBOW는 문맥 정보의 단어들을 사용하지 않고 문서 벡터

만으로 문서에 나오는 단어를 임의로 뽑아내어 예측하는 방식이다[10].

[11]은 Doc2Vec 모델을 영어 위키피디아 문서에 적용하여 ‘Lady Gaga’ 문서와 동일한 문서 벡터 내 연관 있는 문서를 찾았고, 그 결과 동시대의 팝 여가수들을 뽑아냈다. 또, [11]에서는 Doc2Vec 단어 임베딩 모델을 이용하여 쿠텐베르크 사이트에서 제공하는 자료 총 433개의 텍스트를 대상으로 텍스트 장르의 문맥을 학습하고 장르를 예측하였다. 이 연구에서는 텍스트 전반에 고르게 나타나는 장르적 특성을 Doc2Vec 임베딩 기법을 통해 균일하게 추출해 내는 결과를 가져왔다. Doc2Vec이 문서 분류에 효과적이라는 것은 [12]에서도 나타난다. [12]에서는 Word2Vec을 활용한 CNN 모델 기반에 Doc2Vec을 함께 활용하여 한국어 신문 기사를 분류하였고 그 결과 분류 성능이 향상됨을 보여주었다.

본 연구에서는 일반 용어가 사용된 질의문과 연관 관계에 있는 판례를 연결하기 위해 전문 용어를 포함하는 전문가 답변을 매개로 한다. 질의문에 해당하는 답변과 판례를 Doc2Vec에 적용하고 유사도를 기반으로 답변과 관련 있는 판례를 자동으로 찾아내어 질의문과 연결한다.

3. 제안 방법

본 논문에서는 Doc2Vec 문서 임베딩 모델을 사용하여 질의문과 판례를 효과적으로 연결한다. 먼저 질의문, 답변, 판례를 수집하고 전처리한 후, 전문가 답변과 판례 데이터에 Doc2Vec을 적용한다. 이때 질의문 데이터가 아닌 답변 데이터에 모델을 적용한 것은 일반 사용자의 질의문에는 법률용어가 사용되지 않아서 전문용어가 사용된 전문가 답변을 벡터화할 필요성이 있기 때문이다. 그리고 벡터 공간 모델(Vector Space Model)에서 벡터화된 전문가 답변과 판례 문서의 유사도를 계산하여 답변과 가장 유사한 판례를 뽑아내도록 한다. 그 후 답변을 매개로 답변에 해당하는 질의문과 판례를 연결한다. 연구 절차를 도식화하면 그림 1과 같다.

3. 1. 데이터 수집 및 전처리

본 연구에서 대상으로 하는 데이터 중 전문가 답변 문서는 포털 사이트 ‘네이버’ 지식인 Q&A 사회·정치 분야의 법, 법률 카테고리 중 질의문 수가 상대적으로 많은 ‘교통 사고, 위반’ 카테고리 내 답변과, 온라인 Q&A서비스인 ‘아하’ 페이지 내 법률 카테고리에서 수집되었다. 수집된 답변 데이터는 총 1,540건이다. 전문가 답변 문서와 연결한 판례 데이터는 ‘법률종합정보’ 사이트¹⁾를 이용하여, ‘교통사고’ 키워드를 검색하여 출력되는 판례 총 2,264건이다. 그리고 수집한 데이터를

임베딩하기 위해 전처리 과정을 통해 특수문자를 비롯한 불필요한 문자를 제거하여 NLTK 패키지를 이용하고 문장을 토큰화하였다.

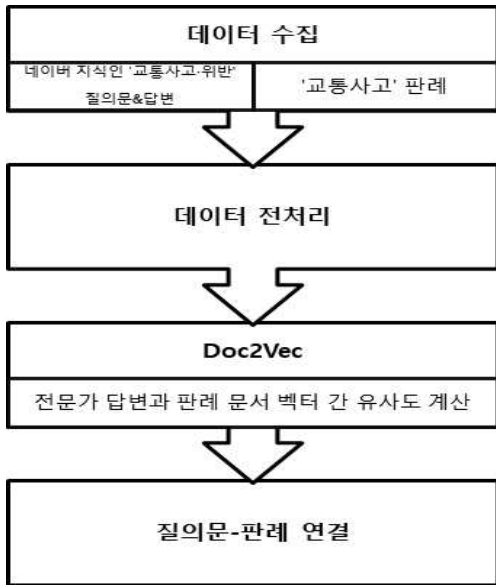


그림 1 질의문-판례 연결 과정

3. 2. Doc2Vec 모델 학습

본 연구에서는 Doc2Vec 모델을 ‘교통사고’와 관련한 질의문에 해당하는 전문가 답변과 판례에 적용하였다. Doc2Vec 훈련 알고리즘 중 PV-DM(Distributed Memory)을 적용해서 모델을 생성하였다. 모델에 사용한 패키지는 scikit-learn 이며, gensim 임베딩 패키지를 사용하였다. 모델의 parameter는 다음과 같이 설정하였다.

표 1 Doc2Vec 모델 parameter

Option	Value
DM	1
Vec_size	100
alpha	0.025
min_alpha	0.0025
min_count	10
epoch	100
window	10

위 표1과 같이 DM은 1로 문서 벡터와 단어 벡터 모두를

사용하는 Doc2Vec 모델의 PV-DM 알고리즘을 사용하였다. Window는 10으로 설정하여 훈련 시 10개의 단어를 고려하도록 하였고, 데이터에서 등장하는 단어의 최소빈도수는 10으로 설정하였다. 훈련 횟수는 100으로 하여 모델을 생성하도록 하였다.

3. 3. 질의문-판례 연결

질의문과 판례를 연결하는 과정은 다음과 같다. 먼저 답변 문서와 판례 문서에 3. 2절의 DM 모델을 적용하여 문서를 임베딩하여 문서를 벡터화한다. 그리고 답변 문서와 판례 문서 간 벡터값을 이용하여 두 문서의 유사도를 계산한다. 이때 유사도는 gensim의 내장 함수인 most.similar()를 이용하여 계산하였다. most.similar() 함수를 통해 특정 답변 문서와 가장 유사한 상위 5개의 답변 문서를 추출하여, 특정 답변과 연관성이 높은 답변을 비슷한 주제로 분류할 수 있도록 하였다. 또한, 벡터화한 답변 문서와 유사도가 높은 판례가 추출될 경우 이를 관련 판례로 제시한다. 특정 답변에 대한 유사 답변과 관련 판례를 제시할 시 해당 답변과 판례는 일대다, 다대일 또는 다대다의 대응관계를 가지도록 한다. 즉 하나 또는 여러 개의 답변과 대응하는 유사도가 높은 여러 답변 혹은 판례를 제안하는 것이다. 이는 판례가 하나의 법률 주제만 다루는 것이 아니라 둘 이상의 법률에 관련되기 때문이다. 마지막으로 그림 2와 같이 유사도가 높은 답변은 비슷한 주제의 답변끼리 분류할 수 있도록 하고, 도출된 관련 판례문은 최종적으로 질의문과 연결하여 준다.

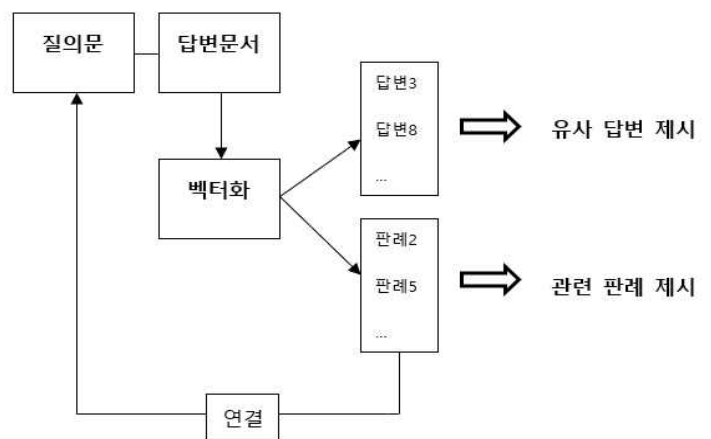


그림 2 관련 답변 제시 및 질의문-판례 연결 과정

4. 질의문-판례 연결 결과

교통사고와 관련한 답변 총 1,540건과 판례 문서 총

1) <https://glaw.scourt.go.kr/wsjo/intsrch/sjo022.do>

2,264건에 Doc2Vec 문서 임베딩을 적용하여 두 문서의 벡터 값을 이용하여 문서 간 유사도를 계산하였다. 그 결과 아래 표 2와 같이 유사 답변들을 제안하여 비슷한 주제의 답변끼리 분류가 가능하도록 하였다. 또한 표 3과 같이 특정 답변과 유사도가 높은 판례를 제안하여 질의문과 판례를 연결할 수 있었다. 본 논문에서는 연결된 질의문과 판례 결과 중 일부를 나타내었다.

표 2 유사 답변 제안 결과

질의문	답변	유사 답변	유사도
신호위반 사고 형사합의금 전치 2주 나왔는데 형사합의금 받을 수 있나요. 백대영인데. 2주라 그냥 벌금 맞을 수도 있나요. 가해자가.	귀하의 사고 경위가 가해자의 신호 위반으로 발생한 경우라면 교통사고처리특별법상 12대 중과실 사고에 해당이 되어 가해자는 형사처벌의 대상이 되고 따라서 가해자가 형사합의를 할 생각이 있다면 귀하께서는 가해자와 형사합의를 할 수 있습니다. ... 전치 1주당 50만원 ~ 70만원 정도라고들 하므로 위 금액 범위 내에서 적절한 금액으로 형사합의를 하실 필요가 있습니다.	귀하의 사고 경위가 신호등 있는 횡단보도에서 보행자 신호가 초록분일 때 횡단하다 차에 치인 경우라면 교통사고처리특별법상 12대 중과실 사고에 해당이 되어 가해자는 형사처벌의 대상이 되고 따라서 가해자가 형사합의를 할 생각이 있다면 귀하께서는 가해자와 형사합의를 할 수 있습니다. ...	0.95
		귀하의 사고 경위가 가해자가 적임보행만 가입되어 있고 또한 병소나라면 가해자는 형사처벌의 대상이 되고 따라서 가해자가 형사합의를 할 생각이 있다면 귀하께서는 가해자와 형사합의를 할 수 있습니다.	0.94
제가 3일전 교통사고가났습니다. 저는 1차선으로 가고 있는도중 상대방이 2차선에서 1차선으로 들어오려다가 제차 뒷부분을 들이받았습니다. 일단 보험회사 다 불러서 사진찍고 블랙박스 확인했습니다. 상대방이 100프로 과실이 나오더군요. 그래서 저는 일단 차 수리맡기고 제가 일을다녀서. 한의원에 통원치료 받고있습니다. 이런부분은 첫째 합의를 언제쯤 해주는것이좋은건지. 두번째 합의할때 합의금을 열	교통사고로 인하여 피해자분께서 입으신 손해가 특별히 장해가 남지 않는다면 가해자의 보험회사로부터 받을 수 있는 손해배상 항목은 입원기간 동안의 휴업손해, 퇴원 후 통원치료시 1일당 8,000원의 교통비 및 약관 소정의 상해급수에 따른 위자료(15-30만원 정도) 입니다...	항목 중 휴업손해는 입원기간 동안 실제 장해율과 상관없이 장해율을 100% 인정하여 사고 당시 피해자분께서 얻고 있었던 소득을 기준으로 보험회사의 약관상 85%를 보상해 주는 항목입니다...	0.96
		휴업손해는 보험회사의 약관상이나 소송시 법원의 기준에 의해서나 모두 오로지 병원에 입원한 기간 동안에만 보상이 되는 항목이기 때문에 입원 치료를 하지 않는다면 비록 통원치료를 받는 동안 일을 하지 못하여 소득의 상실손해가 발생하였다고 하더라도 휴업손해는 보상	0.95

마정도 받아야 적합할지.		받지 못하게 됩니다. 귀하의 경우에는 사고로 인하여 입은 상해에 대하여 입원을 하지 않고 통원치료만을 받으시는 듯 한데 그렇다면 휴업손해를 보상 받지 못하고 통원치료시 1일당 8,000원의 교통비 및 약관 소정의 상해급수에 따른 위자료(15-30만원 정도)만 보상 받게 됩니다.	
만취상태에서 운전했고 중앙분리대들 들이박고 정차했습니다. 혈중알콜농도는 체혈을했고 수치가 얼마나 나올지 모릅니다만 만취상태여서 0.2 이상은 나올거같은데 초벌인 상태이며 벌금이 얼마 정도 나올까 궁금합니다.	벌금액수는 최소 천만원이상에서 결정됩니다. 1500만원 정도 예상되나 가중감경사유 고려해 벌금액은 1000만원 이상입니다.	귀하의 경우 사고에 대한 책임은 없고 음주운전에 대한 처벌과 행정처분외 문제만 남습니다. 면허정지와 몇백만원 벌금 수준일 것입니다.	0.88

위 표 2와 같이 특정 답변과 유사 답변은 서로 의미적으로 93% 이상의 유사도를 가지며, 비슷한 주제의 답변끼리 잘 분류되어 나타나고 있음을 알 수 있다. 반면, 표 3과 같이 답변에 해당하는 판례를 도출해 낸 결과 실제 유사도는 평균 75% 정도로 나타났다.

표 3 판례 제안 결과

질의문	답변	판례	유사도
자동차사고시 민사법을 처벌 받는데 민사법이 아닌 교통사고특별법을 받나요?	교통사고가 발생할 시 사고 초기부터 가해자의 보험회사는 지불보증을 하여 피해자와 최종 합의시까지 병원에 치료비를 직접 지급을 하고 최종 합의시에 피해자의 과실비율에 따른 치료비 정산을 하게 되므로 피해자분께서는 비급여 항목을 제외하고 최종 합의시까지는 치료비를 병원에 지급할 필요가 없습니다. ...	대법원 1987. 11. 10. 선고 87도1727 판결[업무상과실치상] [집 3 5 (3) 형 , 7 5 7 ; 공 1988.1.1.(815),12 2] 【판시사항】 교통사고처리특별법에서의 교통사고의 의미	0.77
		대법원 1988. 5. 24. 선고 88도255 판결[교통사고처리특별법위반(인정된 죄명:업무상과실치사), 도로교통법위반] [집 3 6 (2) 형 , 3 0 4 ; 공 1988.7.1.(827),10 09] 【판시사항】 가. 교통사고처리특별법의 교통사고에 도로가 아닌 곳(공장내작업장)	0.79

		에서의 사고도 포함하는지 여부(적극) 나. 도로교통법 제 109조 제1호의 처벌규정의 적용범위	
교통사고로 전치4주가 나왔는데 과실은 10:0입니다. 입원에 대한 보상은 입원만 해당되나요?? 통원치료는 아닌가요??	교통사고로 인하여 피해자분께서 입으신 상해가 특별히 장애가 남지 않는다면 가해자의 보험회사로부터 받을 수 있는 손해배상 항목은 입원기간 동안의 휴업손해, 퇴원 후 통원치료시 1일당 8,000원의 교통비 및 약관 소정의 상해급수에 따른 위자료(15-30만원 정도)입니다.	대법원 2014. 2. 27. 선고 2013도15499 판결[도로교통법위반][미간행] 【판시사항】 교통사고를 낸 차의 운전자 등에게 도로교통법 제54조 제2항 본문에서 정한 신고의무가 있는 경우 및 위 조항 단서의 해석	0.74
	귀하의 경우 사고로 인하여 디스크의 진단을 받으셨는데 그렇다면 반드시 사고 후 약 6-8개월이 경과한 후 병원(정형외과)에서 장해진단서를 발급 받아 보험회사와 보상절차를 진행하셔야 합니다. 귀하께서는 40%의 외상기여도가 있다는 의사의 소견서를 받으셨다고 하셨는데 장해진단서를 발급 받아 장해기간, 기왕증기여도가 반영된 장애율을 확인하셔야 합니다.		0.73

판례는 전문가 답변과 텍스트 길이, 구조, 문체 등이 다르며 이는 서로 다른 답변 문서와 판례 문서에서 텍스트 자체의 장르적 특성이 유사도에 영향을 미치는 것을 알 수 있었다. 따라서 특정 주제의 질의문에 해당하는 답변은 비슷한 주제를 가지는 답변과 높은 유사도를 가지며 도출되었지만, 여러 도메인의 법률과 관련한 사건이 복합적으로 다루어지는 판례와는 유사도가 비교적 낮게 계산되었다. 위의 표 2, 3과 같이 샘플 결과를 통해 본 논문에서 제안한 방법론의 효과를 나타내었다.

마지막으로, 별도의 소규모 테스트 평가셋을 구축하여 판례 검색 정확도를 평가하였다. 테스트 평가셋은 60개의 답변 문서로 구성되어 있으며, 답변과 판례가 1:1, 1:다, 다:다의 관계를 가질 수 있도록 복수개의 결과를 허용하였다. 구축한 평가셋을 통해 판례 검색 정확도를 평가한 결과, 71.7%의 정확도(Accuracy)를 보였다.

표 4 판례 검색 정확도 평가

판례 검색 테스트 셋	유사 판례 제시	Accuracy
60	43	71.7

5. 결론 및 향후 연구 방향

본 논문에서는 법률적 도움이 필요한 사람들이 일반용어를 사용한 질의문과 연관 관계에 있는 판례를 자동으로 연결하기 위해 Doc2Vec 문서 임베딩 기법을 사용하였다. 질의문에는 전문 용어가 나타나지 않기 때문에 법률 용어가 나타나는 전문가 답변을 이용하여 답변과 판례 문서를 임베딩하여 문서 간의 유사도를 계산하였다. 그 결과 전문가 답변이 93% 이상으로 유사한 답변을 도출하였고, 답변과 대응하는 판례는 약 75%의 유사도를 가지며 질의문과 연결되었다. 또한, 방법론의 타당성을 확보하기 위해 판례 검색 정확도를 평가하였고, 그 결과 71.7%의 정확도를 보였다.

이처럼 문서 간 유사도를 기반으로 한 유사 답변 자동 분류 및 제시는 사용자 입장에서 소송 주체에 대한 이해를 돕는 역할을 할 것이다. 또한 답변과 관련한 자동 판례문 제시는 질의문과 답변에 대한 법적 근거를 제시함으로써 일반인 법률 정보 검색에 기여한다는 점에서 의의가 있다. 향후 연구로는 더 많은 데이터를 확보하여 다양한 머신러닝 연구에 사용되고 있는 BERT를 판례 검색에 적용할 예정이다.

참고문헌

- [1] 한국경제, “<https://www.hankyung.com/society/article/2018013056021>”
- [2] 정훈, 이무훈, 도하나, 최의인, “사용자 질의 의미 해석을 위한 선호도 기반 검색 기술”, 디지털융복합연구, 제11권, 제2호, pp.271-277, 2013.
- [3] Quoc Le, Tomas Mikolov "Distributed Representations of Sentences and Documents" ICML'14 Proceedings of the 31st International Conference on International Conference on Machine Learning Volume 32 pp.1188-1196(9) Beijing, China June 2014.
- [4] 장인호, “온톨로지 기반 법률검색시스템의 구축 및 평가에 관한 연구”, 성균관대학교 대학원 박사학위논문, 2010.
- [5] 백종범, 이수원, “웹 마이닝을 활용한 법령정보검색 지원 시스템”, 정보과학회논문지: 소프트웨어 및 응용, 제40권, 제7호, pp.395-404, 2013.
- [6] 강예지, “용어 대응 기반 질의문과 판례 연결 방안 연구”, 연세대학교 대학원 석사학위논문, 2020.
- [7] Harris, Z. “Distributional structure”, Word,

- 10 (23), pp.146-162, 1954.
- [8] Firth, J.R. "A synopsis of linguistic theory 1930-1955", *Studies in Linguistic Analysis*, pp.1-32, 1957
- [9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, "Distributed Representations of Words and Phrases and their Compositionality" *NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* pp.3111-3119(9) Lake Tahoe, Nevada December 2013.
- [10] Quoc Le, Tomas Mikolov "Distributed Representations of Sentences and Documents" *ICML'14 Proceedings of the 31st International Conference on International Conference on Machine Learning Volume 32* pp.1188-1196(9) Beijing, China June 2014.
- [11] 김동성, "Doc2Vec 단어 임베딩 언어 모델을 활용한 텍스트 장르 구분", *언어와 정보*, 제23권, 제2호, pp.23-43, 2019.
- [12] 김도우, 구명완, "Doc2Vec과 Word2Vec을 활용한 Concolutional Neural Network 기반 한국어 신문 기사 분류", *정보과학회논문지*, 제44권, 제7호, pp.742-747, 2017.