

안전기준의 검색과 분석을 위한 기계독해 기반 질의응답 시스템

김민호[†], 조상현^{**}, 박덕근^{***}, 권혁철^{****}

Machine Reading Comprehension-based Question and Answering System for Search and Analysis of Safety Standards

Minho Kim[†], Sanghyun Cho^{**}, Dugkeun Park^{***}, Hyuk-Chul Kwon^{****}

ABSTRACT

If various unreasonable safety standards are preemptively and effectively readjusted, the risk of accidents can be reduced. In this paper, we proposed a machine reading comprehension-based safety standard Q&A system to secure supporting technology for effective search and analysis of safety standards for integrated and systematic management of safety standards. The proposed model finds documents related to safety standard questions in the various laws and regulations, and then divides these documents into provisions. Only those provisions that are likely to contain the answer to the question are selected, and then the BERT-based machine reading comprehension model is used to find answers to questions related to safety standards. When the proposed safety standard Q&A system is applied to KorQuAD dataset, the performance of EM 40.42% and F1 55.34% are shown.

Key words: Q&A System, Machine Reading Comprehension, BERT, KorQuAd, Safety Standard

1. 서 론

안전사고 발생의 원인은 크게 환경적 원인과 제도적 원인으로 구분할 수 있다. 전자는 기술적, 관리적, 경제적 문제 등으로 인하여 발생하는 것으로서 충분히 예방한다더라도 사고 발생 당시의 여러 상황에 따라 재발할 수 있다. 반면 후자는 잘못된 법령 및 제도로 인해 발생하는 것으로서 해당 법령 및 제도의 개정을 통해 선제적으로 예방할 수 있다[1]. 예를 들

어 허베이 스피리트호 유류오염 사고를 계기로 유조선 이중선체 의무화 도입이 조기 실시되었고, 구제역 발생을 계기로 축산업 허가제가 도입되었다. 또한, 마우나 오션리조트 체육관 붕괴 사고를 계기로 안전관리 사각지대 해소를 위해 2중 시설물 범위가 확대되었으며, 지붕형상을 고려한 습설하중 기준이 추가되었다[2]. 이와 같은 사후적·수동적 관리방식에서 벗어나 각종 불합리한 안전기준을 선제적·능동적으로 정비한다면 안전사고 발생 위험을 낮출 수 있을

※ Corresponding Author : Hyuk-Chul Kwon, Address: (609-735) Busan, Rep. Republic of Korea, TEL : +82-51-510-2875, FAX : +82-51-516-6764, E-mail : hckwon@pusan.ac.kr

Receipt date : Jan. 13, 2020, Approval date : Feb. 13, 2020

[†] Software Education Center, Pusan National University (E-mail : karma@pusan.ac.kr)

^{**} Dept. of Electrical and Computer Eng., Graduate School, Pusan National University (E-mail : delosycho@gmail.com)

^{***} National Disaster Management Research Institute (E-mail : drpark@korea.kr)

^{****} Dept. of Electrical and Computer Eng., Graduate School, Pusan National University

※ This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2013-0-00179, Development of Core Technology for Context-aware Deep-Symbolic Hybrid Learning and Construction of Language Resources)

것이다.

안전기준은 재해 예방, 안전사고 방지 등을 위하여 안전을 확보할 수 있도록 적용하여야 할 기술적 기준을 체계화한 것으로 관계 법령에 따라 정의된다. 그러나 안전기준 적용 대상 및 분야가 같더라도 상황에 따라 관계되는 법령, 소관 부처가 달라 안전기준을 통합 관리하기에 한계가 있었다[3]. 이러한 문제점을 인식하고 안전기준을 통합적으로 관리하는 방안에 관한 연구가 어느 정도 진행되었으며, 대부분 연구가 안전기준을 통합 관리하기 위한 시스템의 설계 및 구축방안에 대한 제안에서 그쳤다[2, 4, 5].

본 연구는 안전기준의 통합적·체계적 관리를 위해 안전기준을 효과적으로 검색하고 분석하기 위한 지원기술을 확보하는 데 그 목적이 있다. 예를 들어, 사용자가 '타워크레인의 와이어로프 지지점은 몇 개여야 하는가?'라고 질문을 하였을 때, 관련 안전기준을 정의하고 있는 법조문인 '와이어로프 설치각도는 수평면에서 60도 이내로 하되, 지지점은 4개소 이상으로 하고, 같은 각도로 설치할 것'을 찾은 다음, 해당 법조문에서 정답에 해당하는 부분인 '4개소 이상'을 반환하는 기술이다. 이를 위해 안전기준 관련 질문과 연관성이 높은 법조문을 검색하고, 검색된 법조문 중에서 질문에 해당하는 대답이 있는 위치를 반환하는 기계독해 기반 안전기준 질의응답 시스템을 제안한다.

제안모형은 대규모 범용말뭉치와 법령문서에 의해 사전학습(pre-training)된 BERT 언어모형[6]을 기계독해 학습말뭉치인 KorQuAD 1.0(Korean Question Answering Dataset)[7]을 활용하여 전이학습(transfer learning)한 모형이다. 전이학습(transfer learning)은 특정 환경에서 학습한 모형을 그대로 사용하여 비슷한 환경에서 재학습하는 것으로, 기존 모형을 활용하기 때문에 처음부터 학습하는 것과 비교하여 아주 저렴한 비용으로 고품질의 모형을 획득하는 방법이다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 살펴보고, 3장에서는 데이터의 전처리 및 질의응답 시스템의 전체적인 구조에 대해 상세히 설명한다. 4장에서는 제안모형의 성능 평가결과를 분석하고, 마지막으로 5장에서 향후 연구 방향에 대해 논한다.

2. 관련 연구

2.1 안전기준의 통합적·체계적 관리방안 연구

안전사고 발생에 따른 위험을 줄이기 위해서는 안전기준 및 안전 관련 법·제도에 대해 기존 사후적·수동적 관리방식에서 탈피하고, 각종 불합리한 안전기준을 선제적·능동적으로 정비하는 체계로의 전환이 시급하다. 이에 따라 행정안전부는 안전기준을 총괄·관리하기 위해 안전기준 심의제도, 안전기준 등록 관리시스템 등을 추진 중이다. 특히, 안전사고 발생 시 산발적으로 개편되는 안전기준을 최소화하고 상호 간 통일성을 확보하기 위해 안전기준의 효과적인 분석·관리 지원기술 확보가 필요한 상황이다. 이를 위해 안전기준을 통합적으로 관리하는 방안에 관한 다양한 연구가 진행되었다.

[2]에서는 법령으로부터 안전기준을 추출 및 분석하기 위하여 언어처리기술, 지식베이스 구축 등을 활용한 프로세스 설계방안을 제시하였고, [3]에서는 안전기준을 통합 관리하기 위한 메타데이터 요소를 제안하였다. [4]에서는 안전기준을 분류하고 안전관리를 체계화하기 위해 안전관리 개체관계도 설계를 시도하였고, [5]에서는 한국과 일본의 제품안전 확보체계에 대한 비교분석을 통하여 우리나라 제품안전 확보체계의 개선과제와 발전 방향을 제시하였다. 또한 [8]에서는 안전기준을 DB화하여 관리할 수 있는 안전기준 관리시스템을 설계하였다. 그러나 대부분의 연구가 시스템의 설계 및 구축방안에 대해 제안하는데 그쳤다.

2.2 기계독해 기반 질의응답 시스템

기계독해는 기계가 지문과 질문을 입력받고 해당 지문과 질문을 이용하여 지문 내에서 정답을 찾는 자연어처리의 과제로서 자동 질의응답 시스템의 핵심기술 중 하나이다. 한국어 기계독해를 위한 학습말뭉치로는 KorQuAD v1.0[7]이 있다. KorQuAD v1.0은 영문 표준 질의응답 학습말뭉치인 SQUAD v1.0 [9]와 같은 기준으로 구축하여, 이는 한국어 위키백과 문서에 대해 크라우드소싱(crowdsourcing)을 통해 만든 70,000+개의 질문-답변 쌍을 포함한 대규모 학습말뭉치이다. KorQuAD를 이용한 기계독해 모형은 다양한 구문, 어휘 변형을 이해하고 대상, 인물, 시간, 장소, 방법, 원인 등의 정답을 추론하는 독해

능력을 학습해야 한다[7].

이를 위해 [10]은 사전학습된 BERT에 Global Attention을 적용하고 Multi-level Co-Attention Fusion을 결합한 모델을 제안하였으며, [11]은 사전 학습된 BERT에 SRU(simple recurrent unit)와 질의 응답에 적합한 자질을 추가한 모델을 제안하였다. 그러나 이처럼 KorQuAD와 같은 기계독해 학습말뭉치를 이용한 모형은 주어진 정답이 존재하는 1~2개의 문단을 미리 알려주고, 해당 문단에서만 답을 찾도록 학습되었기 때문에, 전체 문서에서 답을 찾아야 하는 질의응답 시스템에 적용하기에는 한계가 있다.

최근에는 이러한 한계를 극복하고자 전체 문서에서 주어진 질문에 대응되는 답을 찾는 기계독해 기반 질의응답 시스템이 국내외에서 연구되고 있다.

국외의 경우 [12]는 SQuAD를 통해서 학습한 기계독해 모델을 이용하여 문서 검색기에서 검색된 위키 피디아 문서 중에서 정답을 찾는 원거리 감독방법을 제안하였다. [13]은 원거리 감독을 적용했을 때 부정확한 라벨링으로 인한 성능 감소를 줄이기 위해 노이즈가 있는 문서를 제외하는 방법을 제안했다. [14]는 문서 검색기를 통해 추출한 문서를 단락으로 나누고 질문과 연관성이 큰 단락들을 선별하여 문서 독해기의 입력으로 사용하였다. [15]는 추출된 정답을 다른 단락들에서 포함하고 있다면 정답으로 선택될 확률을 높게 하고 각 단락에서 추출된 정답을 다시 단락 및 질문과 인코딩하여 연관성을 계산하여 정답을 재순위화하는 방법을 제안했다.

국내의 경우 [16]은 검색된 문서에 정답의 존재 여부를 판단하는 NIL-Aware를 Reinforced Ranker-Reader에 추가한 모델을 제안했으며, [17]은 NIL-Aware 모델에서 찾을 수 없는 질문에 대해 지식기반 질의응답 시스템에서 정답을 추출하도록 결합하여 성능을 높이는 방법을 제안했다.

본 연구에서는 [12]~[15]에서 제안한 방법을 법령 문서에 적용한 안전기준 질의응답 시스템을 제안한다. 제안모형은 전체 법령문서에서 안전기준 질문과 관련된 법령문서를 찾는 다음, 해당 법령문서를 조문 단위로 나눈다. 그리고 해당 조문 중에서 질문의 답에 해당하는 부분을 담고 있을 가능성이 큰 조문만 선별한 다음, BERT 기반의 기계독해 모형을 이용하여 안전기준과 관련된 질문의 답을 찾아준다.

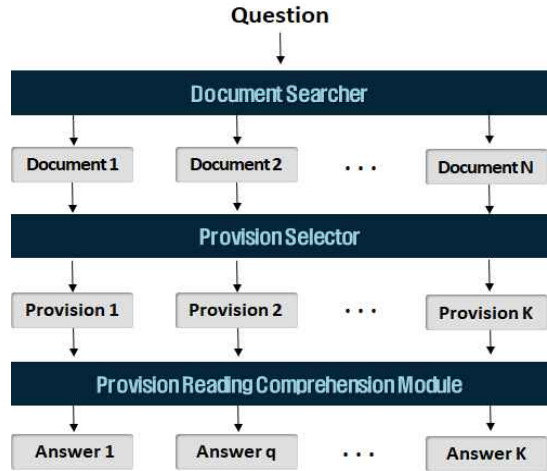


Fig. 1. Structure of safety standard Q & A system.

3. 기계독해 기반 안전기준 질의응답 시스템

본 논문에서 제안하는 기계독해 기반의 안전기준 질의응답 시스템은 Fig. 1과 같이 크게 4가지 부분으로 이루어져 있다: (1) 문서 검색기는 안전기준 질문과 연관성이 큰 법률 문서를 검색한다 (2) 조문 선별기는 인공지능경망 기반의 모형을 이용하여 문서 검색기에서 검색한 다수의 문서에서 질문과 연관성이 큰 조문들을 추출한다 (3) 조문 독해기는 추출된 조문들에서 질문의 응답에 해당하는 부분을 추출한다 (4) 정답 선택기에서는 조문 독해기에서 추출된 여러 개의 후보를 재순위화한다.

3.1 문서 검색기

문서 검색기는 전체 법령문서 중에서 질문과 연관성이 큰 법령문서들을 검색하는 모듈이다. 조문 선별기에서 처리할 문서의 수를 줄임으로써 전체 시스템의 처리 속도를 향상하는 역할을 한다. 전체 법령문서를 색인하여 Fig. 2와 같이 색인어와 문서로 구성된 역색인 파일(inverted index file)을 만들어서 검색에 활용하며, 색인어의 추출을 위한 문서 토큰나이징은 본 연구진이 기개발한 한국어 품사 태거를 사용한다.

사용자의 질문에서 색인어를 추출한 다음에 해당 색인어를 포함한 모든 문서를 찾아 'Okapi BM25'를 적용하여 문서를 순위화한다. 'BM25'는 기본적으로 'TF-IDF 알고리즘'과 유사하지만, 문서의 길이까지 고려한 알고리즘으로서 수식 (1)로 정의된다.

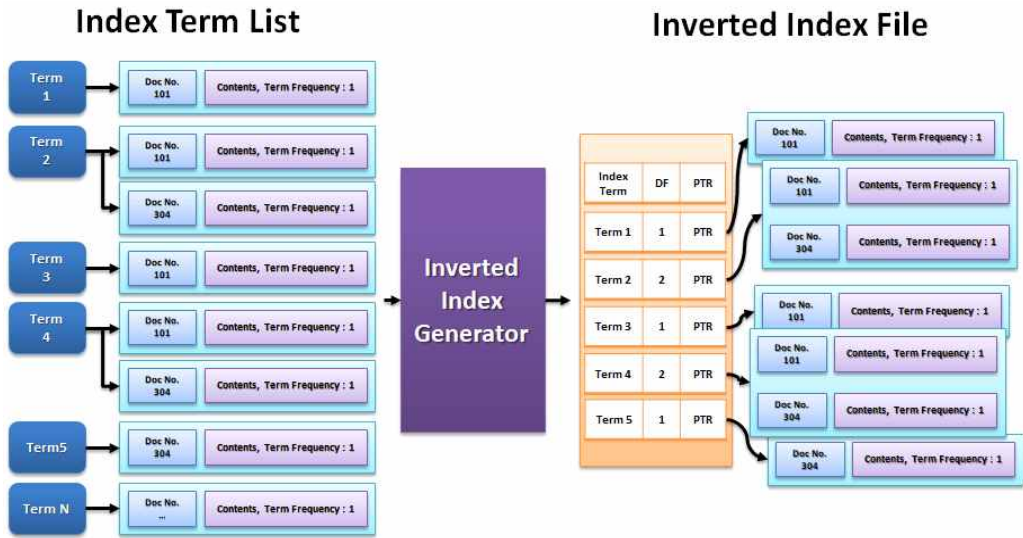


Fig. 2. Structure of inverted index file for document searching.

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{\text{avgdl}})}, \quad (1)$$

사용자 질문 Q 와 검색된 법률문서 D 각각에 수식 (1)을 적용한 다음 내림차순으로 정렬하여 문서를 순위화할 수 있다. 수식 (1)에서 IDF (inverse document frequency)는 기존 IDF 와 유사하지만, TF (term frequency)는 조금 더 복잡하게 계산된다. 이때, $|D|$ 는 현재 문서 D 의 길이, avgdl 은 모든 문서 길이의 평균이다. 나머지 k_1 과 b 는 각각의 단어에 얼마만큼의 가중치를 줄지를 나타낸다. k_1 이 0이 되면 해당 항이 1이 되기 때문에 TF 를 고려하지 않게 되고, 값이 커질수록 해당 단어에 많은 가중치를 주게 된다. b 가 0이 되면 문서의 길이를 고려하지 않게 되고, 1에 가까워질수록 문서의 길이에 긴 가중치를 두어 계산하게 된다.

본 연구에서는 k_1 과 b 의 값을 실험을 통해 결정하였다. 법령문서 검색기의 평가를 위한 평가데이터가 별도로 존재하지 않고, 구축에 큰 비용이 들기 때문에 ‘KorQuad 1.0’ 데이터를 활용하여 검색 성능을 평가하였다. 즉, KorQuAD 1.0의 ‘질문-지문’을 각각 질의어와 해당 질의어와 관련된 문서로 간주하고, 질문이 입력되었을 때 KorQuAD의 전체 지문 중에서 해당 지문을 얼마나 잘 찾았는지를 측정하였다.

Table 1은 ‘KorQuad 1.0’ 데이터를 대상으로 한

검색 성능으로서 질의문과 연관성이 높은 문서 순으로 나열하였을 때, 상위 N 개의 검색 결과 문서가 정답 문서를 포함하고 있는지를 측정한 것이다. k_1 의 값을 0~2 사이의 값으로 0.1씩, b 의 값을 0~1 사이의 값으로 0.1씩 조정하였을 때 k_1 이 1.7, b 는 0.7일 때 가장 좋은 성능을 보였다. 이때 상위 5개의 문서에 정답이 포함된 경우는 55.1%, 상위 100개의 문서에 정답이 포함된 경우는 83.2%, 상위 300개의 문서에 정답이 포함된 경우는 91.3%이다.

3.2 조문 선별기

하나의 질문에 답변하기 위해서 조문 독해기에서 모든 법령문서를 분석하는 것은 많은 처리시간이 필요하다. 조문 선별기는 문서 검색기에서 추출된 상위 100개의 법령문서를 조문 단위로 나누고, 그중에서

Table 1. Performance of document searcher

N -best documents	Accuracy
5	55.1%
10	63.4%
20	68.5%
50	76.8%
100	83.2%
200	88.5%
300	91.3%

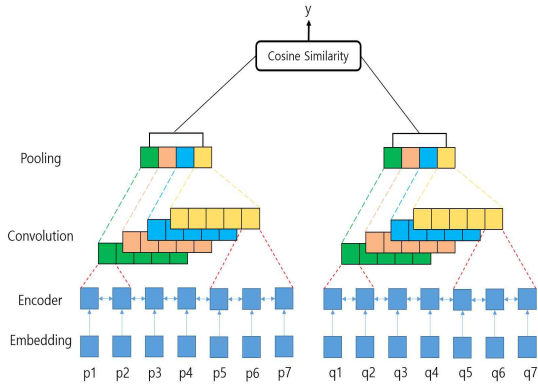


Fig. 3. Structure of provision selector.

질문과 연관성이 큰 조문들을 선별하는 모듈이다. 이때 조문의 길이가 BERT 모형에서 처리 가능한 길이인 512를 넘는 경우 향으로 분리하였다¹⁾.

조문 선별기의 구조는 Fig. 3과 같다. 인코더는 LSTM[18]과 트랜스포머[19]를 이용하며, 컨볼루션과 풀링 레이어는 [20]에서 제안한 방법을 사용하였으며, 이에 대한 전체 수식 (2)와 같다.

$$r_i = \alpha(P_{i,j}, q) * \beta(D_i, q) \quad (2)$$

$D = \{D_1, \dots, D_{100}\}$ 는 문서 검색기를 통해 추출된 상위 100개의 문서이며 $P_i = \{P_{i,1}, \dots, P_{i,i}\}$ 는 검색된 문서를 조문 단위로 나눈 것이며, q 는 질문이다. α 는 조문 선별기의 출력이며 β 는 문서 검색기의 출력이다. r 값이 큰 상위 K 개의 조문만을 문서 독해기의 입력으로 전달한다.

3.3 조문 독해기

조문 독해기는 조문 선별기에서 선별된 조문을 입력으로 받아서 질문에 대한 정답을 추출하는 모듈이다. 3.2절에서 설명한 대로 여러 조문으로 이루어진 법령문서들을 조문 단위로 나누었으며, 길이가 512를 넘는 조문은 조문을 구성하는 향이나 호로 나누었다. 조문 독해기의 기반이 되는 BERT 언어모형은 구글에서 공개한 ‘BERT-Base, Multilingual Cased’

1) 법령의 본칙과 부칙은 조로 구분하며, 하나의 조문에 여러 내용을 규정하거나 세분하여 규정할 때는 조문을 ‘향’이나 ‘호’로 구분한다. ‘향’은 완성된 형식의 문장을 사용하는 반면 ‘호’는 단어나 어절의 형식이나 “... 할 것”과 같은 표현 방식을 사용한다.

를 법령문서에 대해 1세대만 재학습하면 활용하였다. 재학습에 사용된 법령문서는 행정안전부에서 고시한 ‘안전기준 등록(2018.07.20. 기준)’의 모든 안전기준을 포함한 법령문서로서 전체 445개 법령문서(법률, 시행령, 시행규칙, 27,065개 조문)이다.

실제 조문 독해기의 학습을 위해서는 안전기준 질의응답 데이터를 이용하여 전이학습을 통해 BERT 언어모형을 미세조정(fine-tuning)하여야 한다. 그러나 안전기준 질의응답 데이터가 없으므로 본 연구에서는 KorQuAd 1.0 기계독해 학습데이터를 사용하였다. 대신 KorQuAD 1.0 학습말뭉치에 지나치게 의존하는 것을 막기 위해 무작위로 선택된 질문과 관련이 없는 조문을 학습말뭉치에 포함했다. 정답이 없는 무작위로 선택된 문서의 정답 시작 위치와 끝 위치는 [CLS]를 가리키도록 했다. Fig. 4는 조문 선별기의 구조를 나타낸다.

$X_i = [CLS], q_0, q_1, \dots, q_m, [SEP], p_{i,0}, p_{i,1}, \dots, p_{i,m}$ 와 같은 포맷으로 BERT에 입력을 주게 된다. 여기서 $Q = \{q_1, \dots, q_m\}$ 는 질문을 의미하며, $P_i = \{p_{i,0}, \dots, p_{i,m}\}$ 는 정답을 찾기 위한 i 번째 단락을 의미한다. [6]에서 SQuAD 태스크에 적용한 방법과 같이 출력값의 차원 수를 2로 설정한 FFNN(feed-forward neural network)을 통해서 시작과 끝의 확률을 구하였다. 이에 대한 수식은 다음과 같다.

$$h_i = BERT(X_i) \in R^{(n+m+2) \times d} \quad (3)$$

$$h_{i,j}^o = FFNN(h_{i,j}) \in R^{(n+m+2) \times 2} \quad (4)$$

$$y_i^s = softmax(h_i^{o1}) \in R^{n+m+2} \quad (5)$$

$$y_i^e = softmax(h_i^{o2}) \in R^{n+m+2} \quad (6)$$

여기서 수식 (3)의 d 는 BERT의 은닉층의 차원 수를 의미한다.

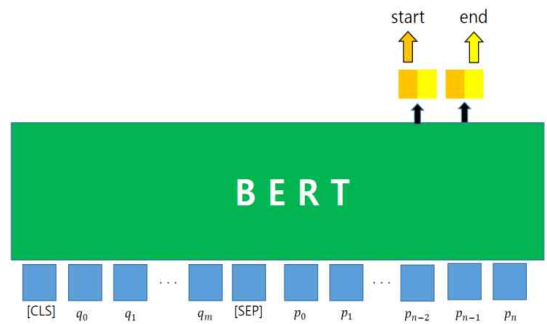


Fig. 4. Structure of BERT-based provision selector.

3.4 정답 선택

조문 선별기에서 추출된 상위 K 개의 조문이 독해기의 입력으로 들어가기 때문에 K 개의 정답이 출력된다. 정답 선택은 이 중에서 가장 확률이 높은 1가지 정답을 선택하는 방법이다. 본 연구에서는 정답 선택을 위해 4개의 모형을 제안하고 각 모형을 실험을 통해 비교하였다.

모형 1은 FFNN에서 출력된 모든 단락의 출력값을 병합하고, 병합된 출력값에 softmax를 이용하여 정답의 시작과 끝일 확률을 계산하는 방법이다. 이에 대한 수식은 다음과 같다.

$$\bar{h}^o = [h_0^o; h_1^o; \dots; h_D^o] \in R^{(n+m+2) \times D \times 2} \quad (7)$$

$$\hat{y}^s = \text{softmax}(\bar{h}^{o1}) \in R^{(n+m+2) \times D} \quad (8)$$

$$\hat{y}^e = \text{softmax}(h^{o2}) \in R^{(n+m+2) \times D} \quad (9)$$

수식 (7)의 [;]는 벡터의 결합을 의미한다.

모형2는 [CLS] 토큰을 정답의 시작과 끝으로 선택할 확률이 가장 낮은 단락의 정답을 선택하는 것이다. 이에 대한 수식은 다음과 같다.

$$\hat{A} = \underset{A_i}{\operatorname{argmax}} p(A_i | P_i, Q) \quad (10)$$

$$p(A_i | P_i, Q) = -(y_{i,0}^s + y_{i,0}^e) \quad (11)$$

여기서 A_i 는 기계 독해기에 단락 P_i 를 입력으로 하였을 때 추출된 정답을 의미한다.

모형3은 정답으로 추출된 단어들을 토큰으로 나누고 빈도수가 높은 토큰을 가지는 정답이 선택될 확률을 높이는 방법이다.

$$p(A_i | P_i, Q) = (2 - y_{i,0}^s - y_{i,0}^e) * \frac{1}{T_i} * \sum_{t=0}^{T_i} \text{freq}(A_{i,t}^t) \quad (12)$$

여기서 A^t 는 토큰으로 나누어진 추출된 정답을 의미한다.

모형4는 계산된 확률값에 a 벡터를 곱해주는 방법이다. 이에 따른 수식은 다음과 같다.

$$a_i = \text{FFNN}(h_{i,0}) \in R \quad (13)$$

$$\bar{y}_{i,j}^s = y_{i,j}^s \circ a_i \quad (14)$$

$$\bar{y}_{i,j}^e = y_{i,j}^e \circ a_i \quad (15)$$

4. 실험 및 결과

4.1 실험 환경

본 논문에서는 제안하는 안전기준 질의응답 시스템은 법령문서에서 안전기준과 관련된 질문의 답을 찾는 기계독해 기반의 질의응답 시스템이다. 그러나 안전기준 질의응답 데이터를 구축하는데는 큰 비용이 들기 때문에 기존에 존재하는 KorQuAD 1.0 학습 말뭉치를 이용하여 해당 모형을 학습하였다. 즉, 위키피디아 문서 전체에서 특정 질문의 답을 찾는 모형을 만든다면, 해당 모형은 법령문서 전체에서도 안전기준과 관련된 질문의 답을 찾을 수 있다고 가정하였다. 이를 위해 다음과 같이 KorQuAD 1.0 기반의 학습말뭉치를 준비하였다.

KorQuAD 1.0은 총 60,407쌍의 질의응답을 포함한 학습 데이터, 5,774쌍의 질의응답을 포함한 개발 데이터, 그리고 3,898쌍의 질의응답을 포함한 평가 데이터로 구성된다. 각 데이터는 질문과 질문에 해당하는 답을 포함하고 있는 위키피디아 문서의 특정 단락(1~2개 문단), 그리고 단락에서 정답에 해당하는 음절의 위치와 정답의 길이로 이루어져 있다.

먼저 KorQuAD의 단락이 포함된 위키피디아 문서를 획득하고자 한국어 위키피디아 덤프 파일(2018년 4월 기준)을 내려받아, 각 위키피디아 문서에서 HTML 태그를 모두 제거하고 평문(plain text)만 추출하였다. 이 중 KorQuAD의 단락 길이보다 작은 100 음절 이하의 길이를 가지는 문서는 제외하고, 총 384,905개의 문서만을 사용하였다. 그런 다음 KorQuAD의 단락과 위키피디아 문서를 비교하여 위키피디아 문서에 정답에 해당하는 음절의 위치를 표시하였다.

조문 선별기의 학습은 KorQuAD 학습/개발 데이터의 질문, 정답, 그리고 문서 검색기의 검색 결과 중 상위 10개를 이용한다. 이는 KorQuAD의 경우 질문의 정답이 존재하는 단락 1개만 데이터에 포함되어 있지만, 실제로 문서 검색기에서 검색된 다른 문서에도 질문에 대한 정답이 포함되어 있을 수 있기 때문이다. 또한, 이전 연구[12]와 같이 만약 질문에 개체명이 있다면 해당 개체명을 포함하지 않는 문서는 모두 제거하였다.

4.2 조문 선별기의 평가

본 연구에서 사용된 트랜스포머 모델의 하이퍼 파라미터는 “은닉층의 차원 수: 128, 어텐션 헤드 수: 4, 트랜스포머 블록 수: 3”로 설정하였으며, LSTM

Table 2. Performance of provision selector (% , KorQuAd Dev Set.)

Model	Results of provision selection			
	$k=20$	$k=40$	$k=80$	$k=100$
Transformer	55.48	62.58	70.32	72.72
LSTM	51.27	59.15	62.29	63.51

모델은 “은닉층의 차원 수: 200, LSTM의 차원 수: 2”로 설정하였다.

문서 검색기에서 검색된 상위 100개의 문서와 질문을 입력하여 조문 선별기에서 출력된 점수를 기반으로 상위 K 개의 단락을 뽑고 K 개의 상위 단락에 정답 단락이 포함되어 있으면 적중한 것으로 보았다. 상위 100개의 문서를 단락으로 나누었을 때 평균 847.2개의 단락으로 나누어졌다.

Table 2는 조문 선별기가 선별한 상위 K 개의 단락에 정답을 포함하고 있는 단락이 존재하는 경우를 측정하는 것이다. 상위 100개의 단락을 선별하였을 경우, 트랜스포머를 사용하여 인코딩하였을 때 더 좋은 성능을 보였다.

4.3 조문 독해와 정답 선택

Table 3은 정답 선택모형에 따른 성능 비교결과이다.

정답 선택 방법에 따른 실험 결과는 다음과 같으며 조문 선별기에서 상위 20개의 단락을 추출했을 때의 결과이다. [CLS] 토큰을 정답의 시작과 끝으로 선택할 확률이 가장 낮은 단락에서 추출된 정답을 선택하는 것이 좋은 성능을 보였으며, 추출된 정답 토큰의 빈도수를 반영했을 때 EM 0.95%, F1 0.96%만큼 향상된 성능을 보였지만 단락의 개수를 40개로 설정하였을 때는 EM 0.98%, F1 1.38%만큼 하락한 성능을 보였다. 이는 추출되는 정답의 개수가 많아지면서 비

Table 3. Performance comparison by choosing the correct answer (% , dev)

Model	$k=20$		$k=40$	
	EM	F1	EM	F1
1	30.80	46.93	32.67	48.14
2	34.61	52.42	39.13	53.20
3	35.09	52.75	34.11	51.37
4	28.55	35.45	30.16	37.58

Table 4. Performance comparison by number of paragraphs (% , dev)

k	EM	F1
20	34.61	52.42
40	39.13	53.20
80	40.42	55.34
120	41.81	55.09

슷한 유형을 가지는 정답이 아닌 토큰들의 개수가 늘어났기 때문에 성능이 하락한 것으로 보인다.

Table 4는 조문 선별 모델에서 추출하는 단락의 개수에 따른 실험 결과이다. 가장 좋은 성능을 보이는 정답 선택 모형 2를 통해 성능을 비교하였다. 조문 선별에서 더 많은 단락을 추출할수록 질문과 관련된 단락이 추출될 확률이 높아지는 것을 보였으며, 더 많은 단락을 입력했을 때 정답 선택 역시 더 좋은 성능을 보였다. 그러나 단락의 개수가 120개일 때는 EM 1.39%가 향상되었지만, F1의 경우 0.25% 하락했다. 이는 정답 선택 모형에서 선택해야 할 정답의 개수가 많아지면서 정답 선택 모형의 정확도가 떨어진 것으로 보인다.

4.4 안전기준 질의응답

Fig. 5는 조문 독해 모형에 질의문과 질의문의 정답이 포함된 지문을 입력으로 주었을 때의 결과이다. ‘타워크레인의 지지점은 몇 개여야 하는가?’라는 질문(a)에 대해서 각 3개의 조항에서 ‘4개소 이상’, ‘4개 이상’, ‘3개소 이상’이라는 응답을 도출한 것을 확인할 수 있다. 이는 질문에 대한 정답을 포함하고 있는 문서가 제공된다면 안전기준 질의응답 시스템이 정답을 잘 찾을 수 있다는 것을 보여준다. 반면에 ‘소형 선박이란 무엇인가?’라는 질문에는 ‘다음 각호의 어느 하나에 해당하는 선박을 말한다’와 같은 답을 찾았다. 이는 조문 독해기의 학습에 사용된 KorQuAD 데이터에 법률문서의 조문과 같은 구조를 가지는 단락이 없기 때문에 이러한 유형의 단락이 학습되지 않았기 때문이다. 또한 (a)와 같은 질문에서는 ‘몇 개’라고 하는 어휘정답유형(lexical answer type; LAT)이 명확하게 드러나는 질문은 기계독해 모형이 정답을 잘 찾지만, (b)와 같이 정의를 묻는 질문이나 예/아니오 질문과 같은 유형은 기계독해 모형이 정답을 잘 찾을 수 없기 때문이다. 이러한 유형의 질문을 처

질문	타워크레인의 지지점은 몇 개여야 하는가?	
	조항	응답
③ 사업주는 타워크레인을 와이어로프로 지지하는 경우 다음 각 호의 사항을 준수하여야 한다.<개정 2013.3.21> 1. 제2항제1호 또는 제2호의 조치를 취할 것 2. 와이어로프를 고정하기 위한 전용 지지프레임을 사용할 것 3. 와이어로프 설치각도는 수평면에서 60도 이내로 하되, 지지점은 4개소 이상으로 하고, 같은 각도로 설치할 것 4. 와이어로프와 그 고정부위는 충분한 강도와 장력을 갖도록 설치하고, 와이어로프를 클립·샤클(shackle) 등의 고정기구를 사용하여 견고하게 고정시켜 풀리지 아니하도록 하며, 사용 중에는 충분한 강도와 장력을 유지하도록 할 것 5. 와이어로프가 가공전선(架空電線)에 근접하지 않도록 할 것		4개소 이상
③ 타워크레인을 와이어로프로 지지하는 경우에는 다음 각 호의 사항을 모두 준수하여야 한다. 1. 와이어로프를 고정하기 위한 전용 지지프레임은 타워크레인 제작사의 설계 및 제작기준에 맞는 자재 및 부품을 사용하여 표준방법으로 설치할 것 2. 와이어로프 설치각도는 수평면에서 60도 이내로 하고, 지지점은 4개 이상으로 하며, 같은 각도로 설치할 것 3. 와이어로프 고정시 텀버클 또는 긴장장치, 클립, 샤클 등은 한국산업규격 제품 또는 한국산업규격이 없는 부품의 경우에는 이에 준하는 규격품을 사용하고, 설치된 긴장장치, 클립 등이 이완되지 아니하도록 하며, 사용 시에도 충분한 강도와 장력을 유지하도록 할 것 4. 작업용 와이어로프와 지지 고정용 와이어로프는 적절한 거리를 유지할 것		4개 이상
③ 타워크레인을 와이어로프로 지지하는 경우에는 다음 각 호의 사항을 모두 준수하여야 한다. 1. 와이어로프를 고정하기 위한 전용 지지프레임을 사용할 것 2. 와이어로프 설치각도는 수평면에서 60도 이내로 하며 지지점은 3개소 이상 등각도로 설치할 것 3. 와이어로프의 고정부위는 충분한 강도와 장력을 갖도록 설치하고, 와이어로프를 클립·샤클 등의 고정기구를 사용하여 견고하게 고정시켜 풀리지 아니하도록 할 것		3개소 이상

(a)

질문	소형선박이란 무엇인가?	
	조항	응답
이 법에서 '소형선박'이란 다음 각 호의 어느 하나에 해당하는 선박을 말한다. 1. 총톤수 20톤 미만인 기선 및 범선 2. 2. 총톤수 100톤 미만인 부선		다음 각 호의 어느 하나에 해당하는 선박을 말한다.
'소형선박'이라 함은 총톤수 25톤 미만의 선박을 말한다.		총톤수 25톤 미만의 선박을 말한다
'소형선박'이라 함은 제27조제1항제2호의 규정에 따른 측정방법으로 측정된 선박길이 12미터 미만인 선박을 말한다.		제27조제1항제2호의 규정에 따른 측정방법으로 측정된 선박길이 12미터 미만인 선박을 말한다

(b)

Fig. 5. Example of the safety standard Q&A system given a short.

리하기 위해서는 [21]에서와 같이 안전기준을 데이터베이스나 온톨로지를 구축하고 자연어 질문을 SQL이나 SPARQL 질의어로 변환하여 검색하는 방법이 필요하다.

5. 결 론

본 논문에서는 안전기준의 통합적·체계적 관리를 위해 안전기준을 효과적으로 검색하고 분석하기 위한 지원기술을 확보하기 위해 기계독해 기반 안전기준 질의응답 시스템을 제안하였다. 제안모형은 전체 법령문서에서 안전기준 질문과 관련된 법령문서를

찾은 다음, 해당 법령문서를 조문 단위로 나눈다. 그리고 해당 조문 중에서 질문의 답에 해당하는 부분을 담고 있을 가능성이 큰 조문만 선별한 다음, BERT 기반의 기계독해 모형을 이용하여 안전기준과 관련된 질문의 답을 찾아준다. 제안한 안전기준 질의응답 시스템을 KorQuAD에 적용했을 때, 개발 데이터에서 EM 40.42%, F1 55.34%의 성능을 보였다.

제안모형은 당장 실용화하기에는 무리가 있으나, 법령문서를 대상으로 하는 질의응답 시스템의 실현 가능성을 보였고, 추후 성능 개선을 통해 안전기준의 통합적·체계적 관리를 위한 지원도구나 혹은 일반인에게 공개하여 안전기준에 대한 접근성을 높일 수

있다.

향후 연구에서는 안전기준 질의응답 데이터를 구축하여 학습에 이용함으로써 실제 법령문서와 안전기준의 특성을 질의응답 시스템에 반영할 예정이다. 또한 법령문서에 포함된 표의 의미를 해석하여 표로 제시되는 안전기준에 대한 질의응답도 가능하게 할 예정이다.

REFERENCE

- [1] National Disaster Management Research Institute, *Development of Application Technology for Safety Standard Analysis based on Accident Case Studies*, NDMI-PR-2018-04-02-02, 2018.
- [2] National Disaster Management Research Institute, *Research on Process Design for Safety Standard Extraction and Analysis*, NDMI-BA-2016-13-01, 2016.
- [3] H. Lee, S. Lim, and D. Park, "Methodology Design for Safety Standard Analyses Based on Structuralization of Attribute Information," *Proceeding of the Conference of the Korea Society of Hazard Mitigation*, pp. 297, 2018.
- [4] H. Lee and Y. Lee, "Effective Safety Management by the Classification of Safety Standard," *Proceeding of the Conference of the Korea Society of Disaster and Security*, pp. 35-42, 2013.
- [5] J. Kim and T.H. Kim, "Comparative Study of Safety Regulations: Focused on the Applied Frameworks of Ensuring Product Safety in Korea and Japan," *Journal of The Korean Society of Disaster Information*, Vol. 10, No. 1, pp. 116-122, 2014.
- [6] J. Devlin, M. Chang, K. Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv Preprint arXiv:1810.04805*, 2018.
- [7] S. Lim, M. Kim, and J. Lee, "KorQuAD: Korean QA Dataset for Machine Comprehension," *Proceeding of the Conference of the Korea Information Science Society*, pp. 539-541, 2018.
- [8] Ministry of Public Safety and Security, *Development of Management System Through Systematization of Safety Standards*, TRKO201500013431, 2014.
- [9] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text," *arXiv Preprint arXiv:1606.05250*, 2016.
- [10] K.H. Park, S.H. Na, Y.S. Choi, and D.S. Chang, "BERT and Multi-level Co-attention Fusion for Machine Reading Comprehension," *Proceeding of the Conference of the Korea Information Science Society*, pp. 643-645, 2019.
- [11] D. Lee, C. Park, C. Lee, S. Park, S. Lim, M. Kim, et al., "Korean Machine Reading Comprehension Using BERT," *Proceeding of the Conference of the Korea Information Science Society*, pp. 643-645, 2019.
- [12] D. Chen, A. Fisch, and J. Wetson, "Reading Wikipedia to Answer Open-domain Questions," *Proceeding of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1870-1879, 2017.
- [13] Y. Lin, H. Ji, Z. Liu, and M. Sun, "Denoising Distantly Supervised Open-domain Question Answering," *Proceeding of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 1736-1745, 2018.
- [14] J. Lee, S. Yun, H. Kim, M. Ko, and J. Kang, "Ranking Paragraphs for Improving Answer Recall in Open-domain Question Answering," *Proceeding of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 565-569, 2018.
- [15] S. Wang, M. Yu, J. Jiang, W. Zhang, X. Guo, S. Chang, et al., "Evidence Aggregation for Answer Re-ranking in Open-domain Question Answering," *arXiv preprint arXiv:1711.05116*, 2017.

- [16] Y.H. Lee, S.H. Na, Y.S. Choi, and D.S. Chang, "NIL-aware Open Domain Question Answering of Korean," *Proceeding of the Conference of the Korea Information Science Society*, pp. 518-520, 2019.
- [17] J. Min, S.H. Na, Y.S. Choi, and D.S. Chang, "A Hybrid Approach of KB-based and IR-based QAs for Open Domain Question Answering," *Proceeding of the Conference of the Korea Information Science Society*, pp. 569-571, 2019.
- [18] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, 1997.
- [19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al., "Attention is All You Need," *Proceeding of the 31st Conference on Neural Information Processing Systems*, pp. 5998-6008, 2017.
- [20] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proceeding of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746-1751, 2014.
- [21] J.I. Kim, "A Natural Language Retrieval System for Entertainment Data ," *Journal of Korea Multimedia Society*, Vol. 18, No. 1, pp. 52-64, 2015.



김민호

2007년 부산대학교 정보컴퓨터공학부 학사
 2009년 부산대학교 컴퓨터공학과 석사
 2011년 부산대학교 전기전자컴퓨터공학과 박사수료

2020년~현재 부산대학교 소프트웨어교육센터 교수
 관심분야: 자연언어처리, 정보검색, 인공지능, 기계학습



조상현

2019년 부산대학교 정보컴퓨터공학부 학사
 2019년~현재 부산대학교 전자전기컴퓨터공학과 석사과정
 관심분야: 자연언어처리, 자동질의응답, 인공지능, 기계학습



박덕근

1990년 고려대학교 지구환경과학과 학사
 1992년 (미) Univ. of Florida 토목공학과 석사
 1998년 (미) Cornell Univ. 토목환경공학과 박사

2012년~2014년 (미) World Bank 선임재난관리전문관
 1998년~현재 행정안전부 국립재난안전연구원 연구관
 관심분야: 재난안전관리, 국제표준, 안전기준



권혁철

1982년 서울대학교 컴퓨터공학과 학사
 1984년 서울대학교 컴퓨터공학과 석사
 1987년 서울대학교 컴퓨터공학과 박사

1992년~1993년 (미)Stanford 대학 CSLI 방문 교수
 1987년~현재 부산대학교 정보컴퓨터공학부, 인지과학협동과정 교수
 관심분야: 인간언어공학, 정보검색, 인공지능