

軍 보고서 등장 문장과 관련 법령 간 비교 시스템 구축 방안 연구*

정지인
연세대학교 산업공학과
(jinny1987@yonsei.ac.kr)

김민태
연세대학교 산업공학과
(iammt@yonsei.ac.kr)

김우주
연세대학교 산업공학과
(wkim@yonsei.ac.kr)

.....

軍에서 방위력개선사업(이하 방위사업)은 매우 투명하고 효율적으로 이루어져야 함에도, 방위사업 관련 법 및 규정의 과도한 다양화로 많은 실무자들이 원활한 방위사업 추진에 어려움을 겪고 있다. 한편, 방위사업 관련 실무자들이 각종 문서에서 다루는 법령 문장은 문장 내에서 표현 하나만 잘못되더라도 심각한 문제를 유발하는 특징을 가지고 있으나, 이를 실시간으로 바로잡기 위한 문장 비교 시스템 구축에 대한 노력은 미미했다. 따라서 본 논문에서는 Siamese Network 기반의 자연어 처리(NLP) 분야 인공 신경망 모델을 이용하여 軍의 방위사업 관련 문서에서 등장할 가능성이 높은 문장과 이와 관련된 법령 조항의 유사도를 비교하여 위법 위험 여부를 판단·분류하고, 그 결과를 사용자에게 인지시켜 주는 ‘軍 보고서 등장 문장과 관련 법령 간 비교 시스템’ 구축 방안을 제안하려고 한다. 직접 제작한 데이터 셋인 母문장(실제 법령에 등장하는 문장)과 子문장(母문장에서 파생시킨 변형 문장) 3,442쌍을 사용하여 다양한 인공 신경망 모델(Bi-LSTM, Self-Attention, D_Bi-LSTM)을 학습시켰으며 1 : 1 문장 유사도 비교 실험을 통해 성능 평가를 수행한 결과, 상당히 높은 정확도로 子문장의 母문장 대비 위법 위험 여부를 분류할 수 있었다. 또한, 모델 학습에 사용한 子문장 데이터는 법령 문장을 일정 규칙에 따라 변형한 형태이기 때문에 母·子문장 데이터만으로 학습시킨 모델이 실제 軍 보고서에 등장하는 문장을 효과적으로 분류한다고 판단하기에는 제한된다는 단점을 보완하기 위해, 실제 軍 보고서에 등장하는 형태에 보다 더 가깝고 母문장과 연관된 새로운 문장 120문장을 추가로 작성하여 모델의 성능을 평가해본 결과, 母·子문장 데이터만으로 학습시킨 모델로도 일정 수준 이상의 성능을 확인 할 수 있었다. 결과적으로 본 연구를 통해 방위사업 관련 軍 보고서에서 등장하는 여러 특정 문장들이 각각 어느 관련 법령의 어느 조항과 가장 유사한지 살펴보고, 해당 조항과의 유사도 비교를 통해 위법 위험 여부를 판단하는 ‘실시간 軍 문서와 관련 법령 간 자동화 비교 시스템’의 구축 가능성을 확인할 수 있었다.

주제어 : 방위력개선사업, 방위사업, 문장 유사도 비교, 자연어 처리, Siamese Network, Self-Attention

.....

논문접수일 : 2020년 8월 11일 논문수정일 : 2020년 9월 6일 게재확정일 : 2020년 9월 20일

원고유형 : 일반논문(급행) 교신저자 : 김우주

1. 서론

국방부는 강한 국방력 건설을 위해 ‘방위력개선사업¹⁾’을 추진하고 있으며, 방위력개선비로 매

년 10조원 이상을 집행하고 있다. 2020년의 방위력개선비는 국방예산 50.2조원 중 33.3%인 16.7조원을 차지하고 있으며, 이는 2019년 대비 8.5%가 증가한 비용으로서 2006년 방위사업청 개청

* 이 논문은 국토교통부의 스마트시티 혁신인재육성사업으로 지원되었습니다.

1) 방위력개선사업 : 군사력을 개선하기 위한 무기체계의 구매 및 신규개발·성능개량 등을 포함한 연구개발과 이에 수반되는 시설의 설치 등을 행하는 사업<방위사업법 3조 1항>

이후 최고 비중임을 미루어볼 때 방위력개선사업(이하 방위사업)의 중요도는 매년 증가한다고 볼 수 있다. 또한, 방위사업은 국민의 생명과 재산뿐만 아니라 국가의 안보와도 직결되는 만큼 전문가들에 의해 매우 투명하고도 효율적으로 이루어져야 함은 분명하다.

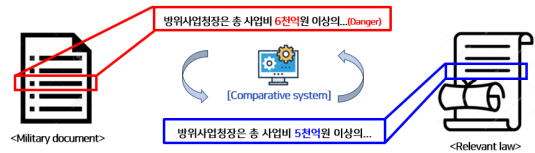
2006년 방위사업청의 개청으로, 기존에 법적 근거 없이 국방획득관리규정에 의해 추진되던 방위사업이 「방위사업법」 제정을 통해 법률에 의해 추진되기 시작했다. 방위사업관리에 적용되는 관련 법 및 규정은 2006년 129개였으나, 2015년에는 174개로 35% 증가되었다. 이러한 상황에서 방위사업관리를 수행하는 담당자는 많은 법과 규정 때문에 업무를 수행하는 데 많은 어려움을 겪고 있으며, 업무를 추진한 이후에 모르는 관련 규정이 있다는 사실을 인식하는 경우도 많은 것으로 알려지고 있다(Kim, 2017, p. 77).

특히, 방위사업 관계 법령(법률, 대통령령, 국방부령) 건수(12건)에 비하여 방사청 소관 행정규칙(훈령, 예규, 고시) 건수가 162건으로 매우 많은 것은 전문적이고 기술적인 특성상 행정규칙을 통해 보다 세부적으로 명시할 필요가 있기 때문이다. 하지만 그에 비례하여 방위사업관리는 더욱 어렵게 되었고 때로는 어떤 법과 규정을 적용해야 하는지도 모르는 상태에서 업무를 담당하는 경우도 있다(Kim, 2017, p. 77).

한편, 법령은 문장 내에서 단어가 하나만 잘못 되더라도 문장의 의미가 크게 바뀌는 아주 예민한 문장이므로 이는 방위사업에서 매우 심각한 문제를 유발할 수 있다. 그럼에도 불구하고 방위사업 추진 간 실무자들은 관련 법령을 직접 손으로 찾아가며 업무를 추진 중이며, 이러한 업무수행 방식은 많은 시간과 노력이 소모될 뿐만 아니라 법령 임의해석 및 오해석(誤解釋)의 위험이

존재함으로써 치명적인 국고의 손실과 직결될 가능성이 있다.

따라서 본 연구에서는 방위사업 관련 실무자가 각종 계획문서 및 보고문서 등 주요 문서 작성 시 참고할 수 있는 자동화 된 ‘軍 보고서 등장문장과 관련 법령 간 비교 시스템’ 구축 방안을 제안하려고 한다. 인공 신경망 모델을 이용한 실험을 통해 軍의 방위사업 문서 내에서 등장할 가능성이 높은 문장(子문장)을 이와 연관된 관련 법령의 해당 조항(母문장)과 1 : 1로 비교해 보았으며 다양한 모델을 통해 실험한 결과, 위에서 언급한 바와 같이 두 문장의 전체적인 구조가 매우 유사하더라도 의미의 차이를 유발하는 작은 차이점도 인식하여 관계 법령상 위법 위험 여부를 상당히 정확하게 분류할 수 있음을 확인했다.



〈Figure 1〉 Comparison system concept

2. 관련 연구

2.1. 문장 표현

Recurrent Neural Networks(RNN)(David E. Rumelhart et al., 1986)은 문장의 각 토큰을 시계열(Sequential) 데이터로 입력받아 처리할 뿐만 아니라 시계열 데이터의 길이에 제한을 받지 않기 때문에, 자연어 처리(NLP) 분야에서 문장을 표현하기 위한 하나의 유용한 인공 신경망 네트워크로 사용되어왔다. 하지만, RNN은 오차 역전

과 과정에서 vanishing gradient problem을 유발함으로써 긴 문장을 처리하는 경우 문장 앞부분의 정보를 충분히 활용하지 못하는 장기 의존성 문제(The problem of Long-Term Dependencies)를 일으킨다. 일상적으로 사용하는 문장보다 상대적으로 길이가 긴 법령 문장의 특성을 고려하면, RNN을 사용한 문장 표현은 적절하지 않음을 알 수 있다.

Long Short-Term Memory(LSTM)(Sepp Hochreiter et al., 1997)은 이러한 문제를 해결한 네트워크 구조로서 RNN의 은닉층 상태(hidden state)(h_t)에 셀 상태(cell state)(c_t)를 추가한 구조이다. 이 셀 상태는 입력 게이트(i_t) · 망각 게이트(f_t) · 출력 게이트(o_t)를 포함하는데, 시계열 데이터에서의 t 시점 셀 상태는 바로 전 시점인 $t-1$ 시점의 셀 상태를 입력으로 받아 각 게이트를 통해 기억해야 할 정보와 잊어야 할 정보를 구분 · 정리하기 때문에 RNN의 장기 의존성 문제를 보완한다. LSTM 셀 상태의 수식은 아래와 같다.

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \\ h_t &= o_t \tanh(c_t) \end{aligned} \quad (1)$$

앞에서 설명한 정방향 RNN 및 LSTM은 과거 시점의 정보를 미래 시점의 정보를 예측하는데 사용할 수 있지만, 반대로 미래 시점의 정보를 과거 시점의 정보를 예측하는데 사용할 수 없는 단점이 있다. 양방향 RNN(Bidirectional RNN)(Schuster Mike et al., 1997)은 이러한 정방향 RNN의 단점을 해결하는데, 기존 정방향 RNN 층에

역방향 RNN 층을 추가함으로써 문맥을 보다 효과적으로 표현할 수 있게 된다. 본 연구에서는 양방향 RNN을 LSTM 구조에 적용한 양방향 LSTM(Bidirectional LSTM)을 문장 표현을 위한 네트워크 구조로 사용한다.

Attention mechanism(Dzmitry Bahdanau et al., 2015)은 번역 등과 같이 시계열 데이터를 시계열 데이터로 변환(Seq2Seq)하는 모델에서 사용하는 네트워크 구조이다. Seq2Seq 모델에서 앞부분의 시계열 데이터를 특정 벡터로 변환하는 부분을 인코더, 변환된 벡터를 다시 시계열 데이터로 변환하는 부분을 디코더라고 하는데, 이때 시계열 데이터의 길이가 길수록 모델의 성능은 나빠진다. 따라서 Attention mechanism은 디코더의 특정 시점 출력 값이 인코더의 전체 시점 중에서 어떤 시점과 가장 연관이 있는가를 평가한 후에 그 부분에 집중하여 디코더의 해당 시점 출력 값을 계산한다.

Self-Attention(Zhouhan Lin et al., 2017)은 이러한 Attention mechanism을 인코더에서만 사용함으로써 문장을 임베딩(문장 전체를 특정 벡터 공간 상의 점으로 표현)하는 기법이다. 최초 임베딩된 문장을 Bidirectional LSTM 층에 통과시킨 결과를 다시 Attention mechanism의 입력으로 넣음으로써, 하나의 문장을 표현할 때 문장에서 더 중요한 역할을 하는 특정 토큰에 더 많은 가중치를 준다. Self-Attention은 가변 길이의 문장을 고정 길이의 문장 표현으로 나타낼 수 있으며, LSTM 은닉층의 모든 정보를 담을 필요 없이 핵심 정보만을 담을 수 있기 때문에 보다 효과적인 문장 표현이 가능하게 한다. 본 연구에서는 母문장 대비 子문장의 위법 위험 여부를 분류함에 있어 子문장에 포함된 특정 토큰의 위험성을 효과적으로 감지하기 위해 Self-Attention을 통한 문장 표현 기법을 사용한다.

Lee et al.(2019)은 질의응답 시스템 내 정답 추출의 성능 향상을 위한 연구를 진행하였는데, 문장의 문법적인 정보를 정확하게 반영하기 위해 문장의 의존 구문 분석 결과를 문장 표현에 포함했다. 한국어의 경우 어순 구조가 자유롭고 문장의 구성 성분 생략이 빈번하기 때문에 의존 구문 분석이 효과적인 한국어 문장 표현에 도움을 준다고 판단하였으며, 의존 구문 분석 결과를 Graph2vec을 통해 문장 임베딩에 추가했다. Bidirectional LSTM-CRF 기반의 정답 추출 모델을 제안하였으며, 한국어 위키백과, 네이버 지식백과, 네이버 뉴스 데이터를 이용하여 실험하였고, 의존 구문 분석 결과를 문장 표현에 포함하는 것이 정답 추출 시스템 성능 향상에 기여한다는 것을 확인했다.

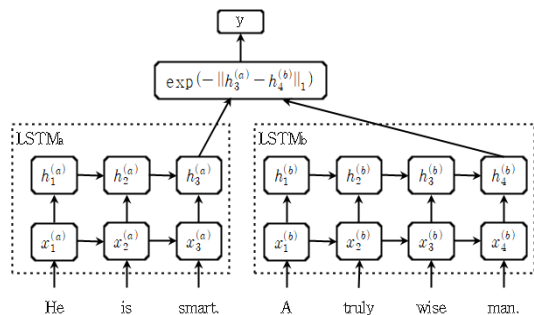
Park et al.(2019)은 Convolution Neural Net (CNN)-LSTM 조합 모델을 이용한 영화리뷰 감성 분석의 분류 정확도를 개선하였는데, 감성 분석에 많이 활용되었던 나이브베이즈(Naive Bayes), 의사결정나무(Decision Tree), kNN, SVM(Support Vector Machine) 등의 기존 기계학습 기법에 대비하여 CNN-LSTM 조합 모델이 IMDB 리뷰 데이터 셋에서 더 높은 성능을 보이는 것을 증명하였으며, 이는 CNN이 문장의 특징을 자동으로 추출하여 병렬 처리한다는 장점과 LSTM이 장기 의존성 문제(The problem of Long-Term Dependencies)를 해결한다는 장점이 복합적으로 작용함으로써 효과적인 문장 표현이 가능하기 때문임을 확인했다.

Lee et al.(2019)은 문장 분류 Task를 수행함에 있어 선택적 단어 제거를 수행한 후 임베딩을 적용하여 문장 분류 정확도를 향상 시키는 방안을 연구하였는데, 텍스트 데이터에서 정보 이득 값이 낮은 단어들을 제거한 후 단어 임베딩을 적용

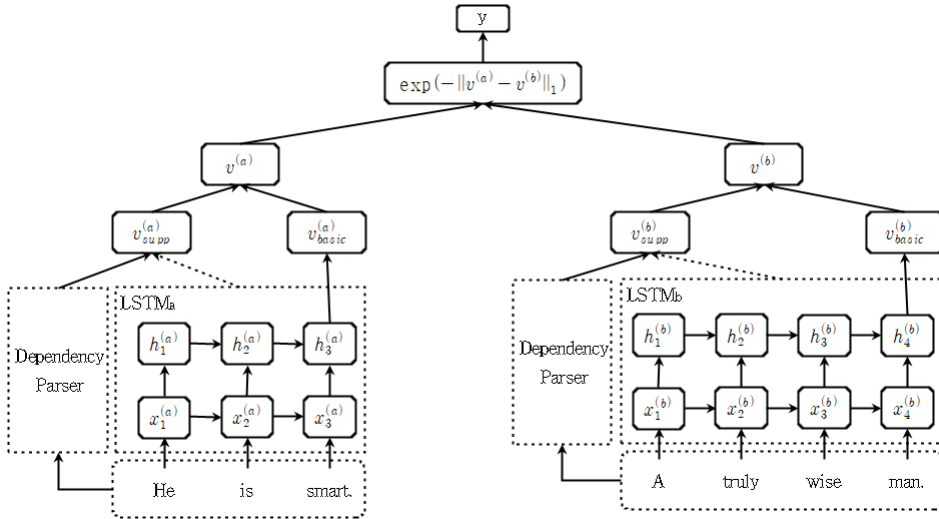
하는 방식과, 정보 이득 값이 낮은 단어와 코사인 유사도가 높은 주변 단어를 함께 텍스트 데이터에서 제거하고 단어 임베딩을 적용하는 방식의 두 가지 방법론을 제안했다. Convolution Neural Net(CNN)과 Attention-Based Bidirectional LSTM을 사용하여 Amazon.com의 ‘Kindle’ 제품에 대한 고객리뷰, IMDB의 영화리뷰, Yelp의 사용자 리뷰 데이터 셋에서 제안된 단어 임베딩 방법의 성능을 확인한 결과, 그렇지 않은 경우에 비해 통계적으로 유의한 차이가 있음을 확인했다.

2.2. 문장 유사도 비교

Siamese Network(Jane Bromley et al., 1994)는 서로 다른 두 개의 입력에 대해 유사도를 비교하는 인공 신경망 네트워크로서, 각 입력(문장)은 인코더를 거쳐 벡터화 된다. 두 개의 입력 벡터는 독립적인 두 개의 인공 신경망 네트워크를 각각 통과하여 두 개의 출력(표현) 벡터로 계산되며, 두 출력 벡터의 거리에 따라 입력 간의 유사도를 판단할 수 있다. 이때, 두 개의 독립적인 인공 신경망 네트워크는 동일한 가중치를 사용한다. Siamese Network는 얼굴 및 음성 인식·이미지 비교·필체 대조 등 최신 기술에 사용된다.



<Figure 2> MaLSTM Model



<Figure 3> D_LSTM Model

MaLSTM(Manhattan LSTM Model)(Jonas Mueller et al., 2016)은 Siamese Network를 응용하였으며, 두 개의 문장 벡터가 각각 통과하는 독립적인 네트워크 구조로는 LSTM을 사용했다. LSTM을 통과한 두 개의 표현 벡터는 Manhattan Distance로 계산되어 유사도를 판단하게 되는데, 이러한 방식은 두 문장의 다양한 특징에서 발생하는 차이를 통합함으로써 두 문장 간의 의미적 유사성을 추론할 수 있으며, 실험에서 사용한 다른 모델(skip-thought RNN, Tree LSTM)에 비해 좋은 성능을 보였다.

D_LSTM(Dependency based LSTM)(Wenhao Zhu et al., 2018)은 MaLSTM이 문장 표현을 만드는데 중요한 요소인 1차 문장 구조(주어, 서술어, 목적어)를 적절하게 다루지 못한다는 단점을 보완하는 구조이다. 하나의 문장 표현(v)은 기존의 LSTM을 사용한 결과인 Basic component(v_{basic})와 의존구문 분석을 사용한 결과인 Supporting component(v_{supp})로 구성된다. 하나의 문장은 의

존구문 분석을 통해 주어·서술어·목적어에는 1, 그 외에는 0으로 라벨링 되고 이렇게 생성된 one-hot 벡터(v_{d_i})는 각 토큰의 LSTM 은닉층 출력 벡터(h_t)와 내적 되어 Supporting component를 생성한다. 이러한 Supporting component는 문장 표현을 최적화하는 역할을 하며, 더 의미 있는 문장 표현 학습이 가능하도록 한다. D_LSTM의 전체 수식은 아래와 같다. 본 연구에서는 子문장의 母문장 대비 위법 위험 여부를 주로 결정짓는 요소가 1차 문장 구조일 것이라는 가정을 하였으며, 문장의 다른 요소보다 1차 문장 구조에 더욱 가중치를 두어 문장을 표현하는 D_LSTM 모델로 비교 실험을 수행했다.

$$\begin{aligned}
 v &= v_{basic} + \alpha \cdot v_{supp} \\
 0 &\leq \alpha \leq 1 \\
 v_{basic} &= h_T \\
 v_{supp} &= \frac{1}{\sum v_{d_i}} \sum_{i=1}^T v_{d_i} \cdot h_i
 \end{aligned}
 \tag{2}$$

Kim et al.(2019)은 입력된 두 문장의 유사도를 측정하기 위한 Siamese CNN-Bidirectional LSTM 모델을 제안한다. 기존 문장 유사도 측정모델의 단어(또는 형태소) 단위의 문장 임베딩이 단어 사전의 크기를 증가시키고 모델의 복잡도를 높이는 문제점이 있다는 점을 지적하였고, 이를 해결하기 위해 문장을 음소 단위로 분해하여 임베딩을 수행했다. 해당 음소를 묶어주기 위해 다양한 필터 사이즈의 1D Convolution Neural Net(CNN)과 LSTM을 결합하였으며, 특히 LSTM 층 이후에 Self-Attention 층을 추가한 모델이 가장 높은 성능을 보였다.

본 논문에서는 Kim et al.(2019)의 인공 신경망 모델을 참고하였으며 네이버 지식인 데이터를 사용한 위 실험과는 달리, 본 논문에서 사용한 법령 문장 및 보고서 등장 문장 데이터는 상당히 정형화되어있는 문장임을 고려하여 음소 단위 임베딩 및 1D CNN 층은 사용하지 않았다. 또한, 본 논문에서는 D_Bi-LSTM 모델을 추가로 구축하여 1차 문장 구조(주어, 서술어, 목적어)가 문장 유사도 비교에 주는 영향을 확인하였으며, 무엇보다도 인공 신경망 모델을 이용한 문장 유사도 비교 연구를 국방 분야로 확장했다는 데 의의가 있다.

2.3. 방위사업 분야 인공 신경망 관련 연구

Han et al.(2011)은 인공 신경망을 이용하여 ‘국방조달 부정당업자 예측모형’을 개발 하였으며, 이를 통해 방위사업 관리자들의 사업 관리 능력 향상과 국방 조달 질서 확립을 도모했다. 사용한 인공 신경망 네트워크는 다중 퍼셉트론(MLP)이며, 국내 276개 군납업체의 데이터(납품 횟수, 납품 금액, 부채비율 등)를 사용했다.

Kim et al.(2011)은 인공 신경망을 이용한 ‘국방 획득사업의 최종사업비 예측 방안’을 연구하였고, 정확하고 신뢰성 있는 최종사업비 예측자료를 바탕으로 사업 관리자의 원활한 의사결정에 도움을 주고자 했다. 다중 퍼셉트론(MLP)을 사용하였으며, 국방 분야 데이터 확보의 제약으로 난수 데이터를 생성하여 실험을 수행했다. 최종사업비를 예측하는데 사용되었던 기존 시스템인 EVMS(Earned Value Management System)와의 성능 비교를 통해, 인공 신경망을 이용한 최종사업비 예측 정확도가 더 우수함을 확인했다.

현재까지 발표된 방위사업 분야 인공 신경망 관련 연구는 본 연구에서 다루는 바와 같이 방위사업 관련 법령과의 위법 위험 여부를 다루고 있지 않으며, 비시계열 데이터를 사용한 예측 모델에 관한 연구가 주를 이루고 있어, 자연어 처리(NLP) 분야를 다루는 본 연구와는 다소 차이가 있다고 할 수 있다.

3. 데이터

3.1. 데이터 형태

실험에 사용한 데이터는 총 3,442쌍의 문장으로서 1 : 1로 비교할 수 있도록 1쌍에는 두 개의 문장이 포함되어있다. 1쌍의 문장은 母문장과 子문장으로 구성되어 있는데, 많은 방위사업 관련 법령 중에서 가장 대표적인 방위사업법·방위사업법 시행령·방위사업법 시행규칙을 대상으로 실무자들이 업무를 하면서 가장 많이 접할 수 있는 주요 조항 83개 조항(母문장)을 선정하였으며, 각 1개 조항(母문장) 당 軍 보고서에서 변형되어 등장할 가능성이 높은 유사 문장(子문장)을

30~50개 작성하여, 총 3,442 쌍의 데이터 셋을 작성했다.

각 쌍의 Label은 ‘母문장 대비 子문장의 위법 위험 여부’로서 1(안전) 또는 0(위험)으로 구분되어 있으며, 모델의 효과적인 학습을 위해 ‘안전’ 및 ‘위험’ 문장의 비율이 약 7 : 3으로 유지되도록 데이터 셋을 작성했다.

3.2. 데이터 생성 규칙

子문장을 만들 때에는 일정한 규칙에 따라 母문장을 변형하였으며, 최대한 각 규칙들이 각 母문장에 균등하게 적용되어 유사한 개수의 子문장을 생성하도록 작업했다.

①번 규칙은 ‘단순화’로서 母문장을 실무 보고서 형태에 맞게 축약·변형하여 子문장을 만든 규칙을 의미한다. 문장의 구성은 단순해졌지만, 내용상 변경된 사항이 없으므로 Label은 1(안전)을 부여했다.

②번 규칙은 ‘주어 순서 변경’이다. 하나의 母문장에 주어가 두 개일 경우, 둘 사이의 순서를 바꾸어 子문장을 만든 형태로서 내용상으로는 변경된 사항이 없으므로 Label은 1(안전)을 부여했다.

③번 규칙은 ‘주어·목적어 변경’으로서 母문장의 주어 또는 목적어가 완전히 다른 주체로 바뀐 子문장을 의미한다. 이는 母문장과 子문장이 문장 형태가 상당히 유사하더라도 주격 대상 또는 목적이 바뀔 경우, 문장이 완전히 다른 의미가 되므로 Label은 0(위험)을 부여했다.

④번 규칙은 ‘수치 표현 변경’이다. 母문장에

수치가 등장할 경우, 단순히 그 수치를 다른 형태로 바꾸어 子문장으로 만든 것을 의미한다. 내용상 변경된 사항이 없으므로 Label은 1(안전)을 부여했다.

⑤번 규칙은 ‘수치 변경’으로서 母문장에 수치가 등장할 경우, 그 수치를 완전히 다른 값으로 바꾸어 子문장으로 만든 것이며, 수치 값의 변경은 매우 중요한 문제를 유발할 수 있으므로 Label은 0(위험)을 부여했다.

⑥번 규칙은 ‘수동형 표현’이며, 母문장의 주격 주체와 목적격 주체를 바꾸어 표현함으로써 문장 형태의 변화를 준 子문장을 의미한다. 의미상으로는 변화가 없으므로 Label은 1(안전)을 부여했다.

⑦번 규칙은 ‘이상·이하 전환’이다. 母문장에서 수치가 등장한 후에 ‘이상’ 또는 ‘이하’라는 표현이 등장할 경우 이 표현을 뒤집음으로써 子문장을 생성했다. 이는 의미상으로 아주 큰 해석의 차이를 유발할 수 있으므로 Label은 0(위험)을 부여했다.

⑧번 규칙은 ‘긍·부정 전환’으로서 母문장의 긍정 또는 부정적 의미를 뒤집어 표현함으로써 子문장을 생성했다. 이 역시 의미상으로 아주 큰 해석의 차이를 유발할 수 있으므로 Label은 0(위험)을 부여했다.

⑨번 규칙은 위에서 설명한 ⑦번 규칙과 ⑧번 규칙을 동시에 적용한 규칙으로서 ‘이상’과 ‘이하’가 뒤바뀌었지만, 긍정과 부정의 의미도 뒤바뀌었으므로 子문장의 전체적인 의미는 母문장과 동일하다. 따라서 Label은 1(안전)을 부여했다.

〈Table 1〉 Rules for editing the original sentence

No	Rule	Label	No	Rule	Label
1	simplification	1	7	switch between more and below	0
2	rearrangement of subject order	1	8	switch between positive and negative	0
3	change subject or object	0	9	simultaneous application of rule 7 and rule 8	1
4	change representation of numerical value	1	10	omission of subject	1
5	change numerical value	0	11	change representation	1
6	Using passive expressions	1	12	change lawful time	0

〈Table 2〉 Data set

No	Original sentence	Edited sentence	Label	edit1	edit2
1	국방부장관 및 방위사업청장은 분석·평가 결과 중 총사업비 오천억원 이상의...	국방부장관은 분석·평가 결과 중 총사업비 오천억원 이상의...	1	1	
		국방부장관 및 방위사업청장은 분석·평가 결과 중 총사업비 육천억원 이상의...	0	1	5
		합동참모의장은 분석·평가 결과 중 총사업비 오천억원 이상의...	0	1	3
.....					
2	국방부장관은 방위력개선사업분야에 관한 중기계획을 대통령의 승인을 얻어 수립한다.	중기계획은 대통령의 승인을 얻어 국방부장관에 의해 수립된다.	1	6	
		중기계획은 국방부장관의 승인을 얻어 방위사업청장에 의해 수립된다.	0	6	3
.....					
3	합동참모의장은 방위력개선사업의 소요에 대하여 무기체계 등의 소요를 결정한다.	합동참모의장은 방위력개선사업의 소요를 결정하지 않는다.	0	1	8
.....					
82	삼십일 이내에 부당이득금등을 방위사업청장이 지정하는 기관에 납부하여야 한다.	부당이득금등은 삼십일 이내에 방위사업청장이 지정하는 기관에 납부되어야 한다.	1	6	
.....					
83	방위사업청장은 업체가..... 일개월 이상 오년 이하의 범위안에서 입찰참가자격을 제한할 수 있다.	국방부장관은 업체가.....입찰참가자격을 제한할 수 있다.	0	3	
		방위사업청장은.....일년 이상 오년 이하의 범위안에서.....	0	5	

⑩번 규칙은 ‘주어 생략’이다. 母문장의 내용이 명확하여 주어 가 굳이 표현되지 않아도 의미 전달에 오해가 발생하지 않거나, 내용 자체에서 주체를 쉽게 연상할 수 있을 경우, 주어를 생략하여 子문장을 생성한 형태이다. 母문장의 내용상 변화가 없으므로 Label은 1(안전)을 부여했다.

⑪번 규칙은 ‘단순 표현 변경’이며, 母문장에서 등장하는 형용사들을 동일한 의미의 다른 형용사로 바꾸어 재표현한 규칙을 의미한다. 따라서 Label은 1(안전)을 부여했다.

마지막 ⑫번 규칙은 ‘적법 시기 변경’으로서 母문장에서 어떤 행위에 대한 기한을 제한할 때, 이 기한 자체를 변경시킴으로써 의미상 심각한 문제를 일으킬 수 있는 것을 의미한다. Label은 0(위험)을 부여했다.

하나의 子문장은 母문장에 하나의 규칙을 적용하여 만들어질 수도 있지만, 다양한 형태의 변화를 위해 대부분 두 개 이상의 규칙이 적용되어 만들어졌다. 데이터 셋 세부내용은 <Table 2>와 같다.

4. 모델

법령 문서에 등장하는 문장은 단어 전체적인 문장 구조가 매우 유사하다하더라도, 단어 차이 하나만으로 행동 주체가 바뀌거나 목적 대상이 바뀌는 등의 심각한 문제를 유발할 수 있다.

합동참모의장은 합동참모회의의 심의를 거쳐 무기체계 등의 소요를 결정한다. (Original Sentence)
 국방부장관은 합동참모회의의 심의를 거쳐 무기체계 등의 소요를 결정한다. (A potentially illegible)

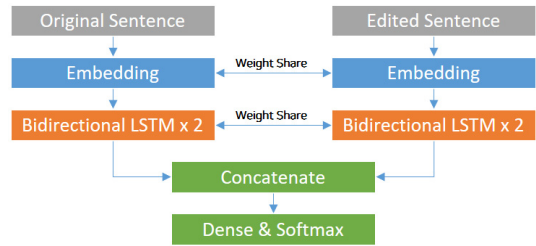
12.8% Inequivalent 87.2% Equivalent

<Figure 4> Characteristic of legal sentence

따라서 실험에 사용되는 모델은 子문장마다

母문장 대비 위법 위험 여부에 가장 결정적으로 영향을 주는 특정 단어에 높은 가중치를 부여함으로써, 母문장과 작은 차이만으로도 분류(위험 또는 안전) Task를 할 수 있어야 한다. 子문장의 위법 위험 여부를 분류할 수 있는 가장 효과적인 기법을 알아보기 위해 아래 세 가지 모델을 구축했다.

4.1. Bidirectional LSTM 모델



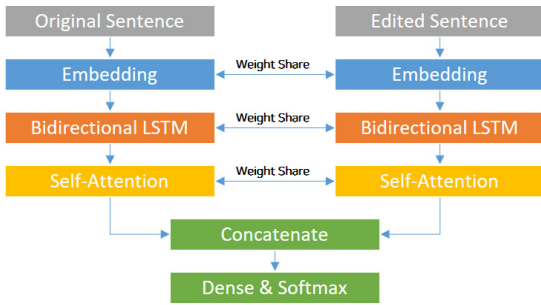
<Figure 5> Bi-LSTM Model

입력된 두 개의 문장(母문장과 子문장)은 임베딩 층을 통과한 후 두 개의 양방향 LSTM 층을 거친다. 양방향 LSTM 층을 통해 문장 표현 벡터를 생성하는 과정에서 먼저 등장한 단어와 뒤에 등장한 단어들 간의 관계를 보다 효과적으로 표현할 수 있을 뿐만 아니라, 이 과정을 두 번 수행함으로써 보다 견고한 문장 표현 벡터를 생성할 수 있다. 子문장의 위법 위험 여부에 직접적으로 영향을 주는 특정 부분을 강조하기 위한 특별한 기법을 적용하지 않아도, 어느 정도의 성능을 보이는지 확인하기 위해 모델을 구축했다.

4.2. Self-Attention 모델

분류 Task를 수행하기 위한 메인 모델로 구축하였으며, 하나의 양방향 LSTM 층을 거쳐 만들

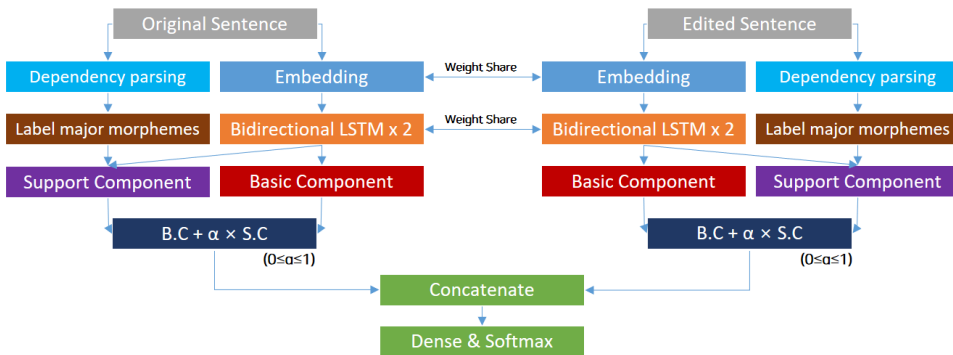
어진 문장 표현 벡터는 Self-Attention 층을 거침으로써 문장 내에서 분류 결과(위험 또는 안전)에 가장 큰 영향을 주는 부분을 집중하여 학습하게 된다. 각 시점의 Bi-LSTM 출력 벡터와 마지막 출력 벡터를 비교하여 가장 영향을 많이 준 시점에 집중하는 context 벡터를 생성하여 분류 학습을 진행한다. 이를 통해 子문장의 위법 위험 여부에 직접적으로 영향을 주는 부분을 강조하여 학습할 수 있다.



〈Figure 6〉 Self-Attention Model

4.3. Dependency based Bi-LSTM 모델

최종 문장 표현 벡터는 Basic Component와 Support Component로 구성된다. 두 개의 양방향



〈Figure 7〉 D_Bi-LSTM Model

LSTM 층을 거쳐 만들어진 문장 표현 벡터는 Basic Component가 되며, 의존구문 분석을 통해 문장의 주요 구성성분(주어 · 서술어 · 목적어)을 강조한 문장 표현 벡터는 Support Component가 된다.

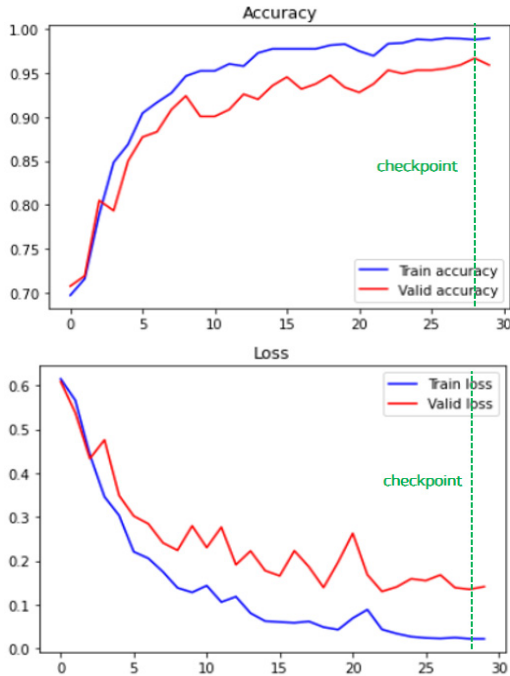
子문장의 주요 구성성분(주어 · 서술어 · 목적어)이 그 외 성분보다 위법 위험 여부에 크게 영향을 준다는 가정 하에 모델을 구축하여 Self-Attention 모델과 성능 비교 실험을 수행했다.

5. 실험 결과

총 3,442쌍의 데이터 중 2,409쌍(70%)을 학습에 사용했고 516쌍(15%)을 검증에, 나머지 517쌍(15%)을 성능 평가에 사용했다. 전체 데이터의 Label 비율(약 7 : 3)을 유지하여 무작위로 데이터를 나누었으며, 평가 방법으로는 정확도 (Accuracy)를 사용했다.

5.1. Self-Attention 모델 학습 결과

단어 임베딩 200차원 · Bi-LSTM 128차원 · 30 에폭(epoch)으로 학습을 수행하였으며, 학습에



〈Figure 8〉 Train result(Self-Attention Model)

폭 중 검증 정확도가 가장 높았던 가중치를 평가에 적용했다. 학습 결과, 학습 정확도 0.98 · 검증 정확도 0.97의 성능을 확인할 수 있었다.

〈Figure 8〉에서 학습된 내용을 모델을 이용하여 517개의 평가 데이터 셋을 분류하였으며, 정확도 0.98 · F1 스코어 0.97의 성능을 보여주었다. 세부 분류 결과는 〈Table 3〉과 같다.

〈Table 3〉 Test result(Self-Attention Model)

Label	Judgment	
	Danger	Safe
Danger	361	4
Safe	5	147

5.2. 모델 비교 실험

모델 비교 실험에 사용한 모델은 위에서 설명

한 Bi-LSTM 모델, Self-Attention 모델, Dependency based Bi-LSTM 모델의 세 가지 모델이며, 전체 데이터 셋의 70%를 학습 · 15%를 검증 · 15%를 평가에 사용했다. 사용자 매개변수(Hyper-parameter)는 세 모델에 동일하게 적용했다.

또한, 학습에 사용된 문장은 법령 문장을 바탕으로 일정 규칙을 적용하여 변경한 형태일 뿐만 아니라 데이터 셋의 크기가 크지 않은 관계로, 위 실험만으로 실제 軍 보고서에서 등장하는 다양한 형태의 母문장 관련 문장들을 효과적으로 분류(위험 또는 안전)한다고 평가하기에는 제한이 있다.

따라서 실제 軍 보고서에 등장하는 형태에 보다 더 가깝고, 母문장과 연관된 새로운 문장 120 문장을 추가로 작성하여 모델의 성능을 비교해 보았다. 이 추가 문장들은 모델 학습 없이, 기존 학습된 모델로 분류 평가를 진행했다. 모델 비교 실험의 결과는 〈Table 4〉와 같다.

〈Table 4〉 Test result(Comparison of three models)

Model \ Data	Test Data set (f1 score)	Military document sentence (f1 score)
Bi-LSTM	0.95	0.74
Self-Attention	0.97	0.81
D_Bi-LSTM	0.97	0.7

5.3. 실험 결과 분석

평가 데이터 셋에 대한 성능은 Self-Attention 모델과 D_Bi-LSTM 모델이 유사하였으나, 실제 軍 보고서 등장 문장에 대한 성능은 Self-Attention 모델이 D_Bi-LSTM 모델보다 높음을 확인할 수 있었다. 이를 미루어보았을 때, 법령

문장의 위법 위험 여부를 결정하는 것은 단순히 문장의 1차 문장 구조(주어·서술어·목적어)만이 아니며, D_Bi-LSTM 모델은 사용자가 지정한 문장 요소(위 실험에서는 주어·서술어·목적어)에만 집중을 하기 때문에 D_Bi-LSTM 모델이 Self-Attention 모델보다 유연성 및 활용도가 떨어짐을 알 수 있다.

한편, 보고서 등장 문장 데이터는 그 수가 많지 않다는 점과 데이터에 Unknown-token(학습에 사용되지 않은 토큰)이 다수 존재한다는 점을 고려시, 큰 의미 부여는 제한됨은 부정할 수 없다. 하지만, 위 실험을 통해 母문장·子문장 데이터 셋을 이용한 모델 학습만으로도 실제 軍 보고서 등장 문장에 대한 분류 성능도 일정 수준 이상 유지한다는 점은 향후 연구의 ‘실시간 軍 문서와 관련 법령 간 자동화 비교 시스템’의 구축을 기대했을 때 상당히 고무적인 결과라고 할 수 있다.

따라서 학습 데이터 셋 및 보고서 등장 문장 데이터의 보강을 통해 충분한 모델 학습이 이루어진다면 더욱더 다양한 軍 보고서 등장 문장의 위법 위험 여부를 분류하는데 긍정적인 효과를 줄 것이다.

6. 결론

본 연구에서는 작은 오기(誤記) 하나로도 심각한 문제를 유발할 수 있는 법령 문장을 효과적으로 학습하여 위법 위험 여부를 판단·분류하고, 이를 사용자에게 인지시켜 주는 다양한 인공지능 실험해보았다.

Siamese Network 구조 기반의 Bi-LSTM 모델, Self-Attention 모델, D_Bi-LSTM 모델을 이용하

여 母문장(실제 법령에 등장하는 문장)과 子문장(母문장에서 파생시킨 변형 문장)간 유사도를 비교한 결과, 子문장의 母문장 대비 위법 위험 여부를 높은 정확도로 분류할 수 있었다.

뿐만 아니라, 母·子문장 데이터를 이용한 학습만으로도 학습 데이터에 존재하지 않는, 실제 軍 보고서에서 사용되는 다양한 문장 형태에 대한 위법 위험 여부 분류도 일정 수준 이상의 성능을 보여주는 것을 확인할 수 있었다.

따라서, 본 실험에서 사용된 학습 데이터보다 많은 양의 母·子문장 데이터와 다양한 형태의 軍 보고서 등장 문장 데이터를 이용하여 보다 충분하게 인공 신경망 모델의 학습을 시도한다면, 추가적인 성능 향상을 기대할 수 있을 것으로 생각한다.

본 연구는 인공 신경망 모델을 이용한 문장 유사도 비교 연구를 국방 및 방위사업 분야로 확장했다는 데 의의가 있는 반면, 현재 연구만으로는 한계점 역시 명확하다. 먼저, 본 논문은 입력된 子문장에 대해 가장 유사하고 연관성이 있다고 판단되는 母문장이 이미 선정되었다는 가정하에 유사도를 비교한 후, 母문장 대비 子문장의 위법 위험 여부를 인공 신경망 모델을 이용하여 분류한다. 따라서, 본 연구의 인공 신경망 모델이 작동하기 이전에 입력된 子문장과 가장 연관성 있는 母문장을 정확하게 선정하기 위한 향후 연구가 필수적이다.

또한, 위에서도 언급되었다시피 학습 데이터의 子문장은 母문장을 일정 규칙에 따라 변형시킨 형태이기 때문에, 실제 軍 보고서에서 등장하는 문장들과는 다소 형태적으로 차이가 있다. 따라서 이를 해소하기 위해서는 Label 작업(관련 법령 대비 위험 또는 안전)이 완료된, 상당량의 실제 軍 보고서 등장 문장의 데이터가 필요하다.

하지만 軍에서 생성되는 데이터는 군사 보안의 특성상 수집이 제한될 뿐만 아니라, 제한적으로 수집이 된다 해도 Label이 0(관련 법령 대비 위험)인 데이터는 존재하지 않을 가능성이 크다. 이러한 어려움 때문에 본 논문에서는 母·子문장 데이터로 학습된 인공 신경망 모델로 실제 軍 보고서 등장 문장을 분류(관련 법령 대비 위험 또는 안전)하는 방안을 제안하였으며, 이를 보다 효과적으로 분류하기 위한 다양한 방안에 대한 추가적인 연구가 필요하다.

결론적으로, 본 연구가 향후 연구로 지속 발전한다면 본 실험의 1 : 1 문장 비교보다 향상된 n : m 문장 비교를 시도해 볼 수 있을 것이며, 방위 사업 관련 軍 보고서에서 등장하는 여러 특정 문장들이 각각 어느 관련 법령의 어느 조항과 가장 유사한지 살펴보고, 해당 조항과의 유사도 비교를 통해 위법 위험 여부를 판단하는, 보다 향상되고 유용한 형태의 ‘실시간 軍 문서와 관련 법령 간 자동화 비교 시스템’ 구축이 가능할 것으로 판단된다.

이러한 ‘실시간 軍 문서와 관련 법령 간 자동화 비교 시스템’은 軍 실무자들이 각종 보고서 작성시 입력되는 다양한 문장들을 실시간으로 처리하여 각 문장들과 밀접하게 연관된 법령들을 찾아주고 해당 법령 대비 위법 위험 여부를 알려줌으로써 관련 업무를 보다 효율적이고 투명하게 수행할 수 있도록 도와줄 것이며, 데이터의 범위만 확대된다면 법령을 다루는 거의 모든 분야에 적용할 수 있는 시스템인 만큼 이러한 모습이 바로 우리 軍과 사회가 미래에 지향해 나갈 스마트한 모습이기를 기대한다.

참고문헌(References)

- Bahdanau, D., K. H. Cho, Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” *ICLR*(2015).
- Bromley, J., I. Guyon, Y. Lecun, E. Sackinger, R. Shah, “Signature Verification using a "Siamese" Time Delay Neural Network,” *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.7, No.4(1994).
- Han, H., S. Choi, “An Artificial Neural Network Approach for the Prediction of Unlawful Company in Defense Procurement,” *Journal of the Military Operations Research Society of Korea*, Vol.37, No.1(2011).
- Hochreiter, S., J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, Vol.9, No.8 (1997).
- Kim, M., H. Han, S. Choi, “A Study on the EAC Estimation of Defense Acquisition Project using Artificial Neural Network,” *Journal of Korea Management Engineers Society*, Vol. 16, No.3(2011).
- Kim, M. T., Y. T. Oh, W. J. Kim, “Sentence Similarity Prediction based on Siamese CNN-Bidirectional LSTM with Self-attention,” *Journal of KIISE*, Vol.46, No.3(2019).
- Kim, S. Y., *Theory and Practice of Defense Acquisition*, Bookorea, 2017.
- Lee, D. K., M. T. Kim, W. J. Kim, “Query-based Answer Extraction using Korean Dependency Parsing,” *Journal of Intelligence and Information Systems*, Vol.25, No.3(2019).
- Lee, M. S., S. W. Yang, H. J. Lee, “Selective Word Embedding for Sentence Classification by Considering Information Gain and Word

- Similarity,” *Journal of Intelligence and Information Systems*, Vol.25, No.4(2019).
- Lin, Z., M. Feng, C. N. Santos, M. Yu, B. Xiang, B. Zhou, Y. Bengio, “A Structured Self-attentive Sentence Embedding,” *ICLR*(2017).
- Mueller, J., A. Thyagarajan, “Siamese recurrent architectures for learning sentence similarity,” *AAAI’16: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*(2016).
- Park, H. Y., K. J. Kim, “Sentiment Analysis of Movie Review Using Integrated CNN-LSTM Model,” *Journal of Intelligence and Information Systems*, Vol.25, No.4(2019).
- Rumelhart, D. E., J. L. McClelland, *Parallel Distributed Processing*, A Bradford Book, Cambridge, 1986.
- Schuster, M., K. K. Paliwal, “Bidirectional Recurrent Neural Networks,” *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, Vol.45, No.11(1997).
- Zhu, W., T. Yao, J. Ni, B. Wei, Z. Lu, “Dependency-based Siamese long short-term memory network for learning sentence representations,” *PLoS One*, Vol.13, No.3(2018).

Abstract

A Study on the Establishment of Comparison System between the Statement of Military Reports and Related Laws

Jiin Jung* · Mintae Kim** · Wooju Kim***

The Ministry of National Defense is pushing for the Defense Acquisition Program to build strong defense capabilities, and it spends more than 10 trillion won annually on defense improvement. As the Defense Acquisition Program is directly related to the security of the nation as well as the lives and property of the people, it must be carried out very transparently and efficiently by experts. However, the excessive diversification of laws and regulations related to the Defense Acquisition Program has made it challenging for many working-level officials to carry out the Defense Acquisition Program smoothly. It is even known that many people realize that there are related regulations that they were unaware of until they push ahead with their work. In addition, the statutory statements related to the Defense Acquisition Program have the tendency to cause serious issues even if only a single expression is wrong within the sentence. Despite this, efforts to establish a sentence comparison system to correct this issue in real time have been minimal. Therefore, this paper tries to propose a “Comparison System between the Statement of Military Reports and Related Laws” implementation plan that uses the Siamese Network-based artificial neural network, a model in the field of natural language processing (NLP), to observe the similarity between sentences that are likely to appear in the Defense Acquisition Program related documents and those from related statutory provisions to determine and classify the risk of illegality and to make users aware of the consequences. Various artificial neural network models (Bi-LSTM, Self-Attention, D_Bi-LSTM) were studied using 3,442 pairs of “Original Sentence”(described in actual statutes) and “Edited Sentence”(edited sentences derived from “Original Sentence”). Among many Defense Acquisition Program related statutes, DEFENSE ACQUISITION PROGRAM ACT, ENFORCEMENT RULE OF THE DEFENSE ACQUISITION PROGRAM ACT, and ENFORCEMENT DECREE OF THE DEFENSE

* Department of Industrial Engineering, Yonsei University

** Department of Industrial Engineering, Yonsei University

*** Corresponding Author: Wooju Kim

Department of Industrial Engineering, Yonsei University

50, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea YONSEI UNIVERSITY College of Engineering D901

Tel: +82-2-2123-7754, E-mail: wkim@yonsei.ac.kr

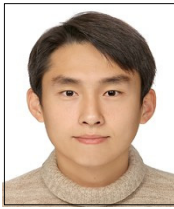
ACQUISITION PROGRAM ACT were selected. Furthermore, “Original Sentence” has the 83 provisions that actually appear in the Act. “Original Sentence” has the main 83 clauses most accessible to working-level officials in their work. “Edited Sentence” is comprised of 30 to 50 similar sentences that are likely to appear modified in the county report for each clause (“Original Sentence”). During the creation of the edited sentences, the original sentences were modified using 12 certain rules, and these sentences were produced in proportion to the number of such rules, as it was the case for the original sentences. After conducting 1 : 1 sentence similarity performance evaluation experiments, it was possible to classify each “Edited Sentence” as legal or illegal with considerable accuracy. In addition, the “Edited Sentence” dataset used to train the neural network models contains a variety of actual statutory statements (“Original Sentence”), which are characterized by the 12 rules. On the other hand, the models are not able to effectively classify other sentences, which appear in actual military reports, when only the “Original Sentence” and “Edited Sentence” dataset have been fed to them. The dataset is not ample enough for the model to recognize other incoming new sentences. Hence, the performance of the model was reassessed by writing an additional 120 new sentences that have better resemblance to those in the actual military report and still have association with the original sentences. Thereafter, we were able to check that the models’ performances surpassed a certain level even when they were trained merely with “Original Sentence” and “Edited Sentence” data. If sufficient model learning is achieved through the improvement and expansion of the full set of learning data with the addition of the actual report appearance sentences, the models will be able to better classify other sentences coming from military reports as legal or illegal. Based on the experimental results, this study confirms the possibility and value of building "Real-Time Automated Comparison System Between Military Documents and Related Laws". The research conducted in this experiment can verify which specific clause, of several that appear in related law clause is most similar to the sentence that appears in the Defense Acquisition Program-related military reports. This helps determine whether the contents in the military report sentences are at the risk of illegality when they are compared with those in the law clauses.

Key Words : Defence Acquisition Program, Sentence Similarity, Natural Language Processing, Siamese Network, Self-Attention

Received : August 11, 2020 Revised : September 6, 2020 Accepted : September 20, 2020

Corresponding Author : Wooju Kim

저 자 소개



정지인

대한민국 해군 소령이며 해군사관학교에서 학사를 취득하고, 군 위탁생으로서 연세대학교 산업공학과에서 석사과정 재학 중이다. 주요 관심 분야는 자연어 처리, 머신 러닝이다. 지능정보시스템학회에서 발표한 바 있다.



김민태

연세대학교 산업공학과에 학사를 취득하고, 연세대학교 산업공학과에서 통합과정 재학 중이다. 주요 관심 분야는 자연어 처리, 추천 시스템, 머신 러닝이다.



김우주

1987년 연세대학교 BBA 과정 학사 학위를 취득하고, 1994년 KAIST 경영과학 박사를 취득하였으며, 현재 연세대학교 산업공학과 교수로 재직 중이다. 관심분야는 시맨틱 웹, 시맨틱 웹 환경의 의사결정지원 시스템, 시맨틱 웹 마이닝, 지식관리 및 인공지능 웹 서비스이다.