

# 지식 그래프와 딥러닝 모델 기반 텍스트와 이미지 데이터를 활용한 자동 표적 인식 방법 연구<sup>☆</sup>

## Automatic Target Recognition Study using Knowledge Graph and Deep Learning Models for Text and Image data

김 중 모<sup>1</sup>                      이 정 빈<sup>1</sup>                      전 호 철<sup>2</sup>                      손 미 예<sup>\*</sup>  
Jongmo Kim                      Jeongbin Lee                      Hocheol Jeon                      Mye Sohn

### 요 약

자동 표적 인식(Automatic Target Recognition, ATR) 기술이 미래전투체계(Future Combat Systems, FCS)의 핵심 기술로 부상하고 있다. 그러나 정보통신(IT) 및 센싱 기술의 발전과 더불어 ATR에 관련이 있는 데이터는 휴민트(HUMINT·인적 정보) 및 시그인트(SIGINT·신호 정보)까지 확장되고 있음에도 불구하고, ATR 연구는 SAR 센서로부터 수집한 이미지, 즉 이민트(IMINT·영상 정보)에 대한 딥러닝 모델 연구가 주를 이룬다. 복잡하고 다변하는 전장 상황에서 이미지 데이터만으로는 높은 수준의 ATR의 정확성과 일반화 성능을 보장하기 어렵다. 본 논문에서는 이미지 및 텍스트 데이터를 동시에 활용할 수 있는 지식 그래프 기반의 ATR 방법을 제안한다. 지식 그래프와 딥러닝 모델 기반의 ATR 방법의 핵심은 ATR 이미지 및 텍스트를 각각의 데이터 특성에 맞게 그래프로 변환하고 이를 지식 그래프에 정렬하여 지식 그래프를 매개로 이질적인 ATR 데이터를 연결하는 것이다. ATR 이미지를 그래프로 변환하기 위해서, 사전 학습된 이미지 객체 인식 모델과 지식 그래프의 어휘를 활용하여 객체 태그를 노드로 구성된 객체-태그 그래프를 이미지로부터 생성한다. 반면, ATR 텍스트는 사전 학습된 언어 모델, TF-IDF, co-occurrence word 그래프 및 지식 그래프의 어휘를 활용하여 ATR에 중요한 핵심 어휘를 노드로 구성된 단어 그래프를 생성한다. 생성된 두 유형의 그래프는 엔터티 얼라이먼트 모델을 활용하여 지식 그래프와 연결됨으로 이미지 및 텍스트로부터의 ATR 수행을 완성한다. 제안된 방법의 우수성을 입증하기 위해 웹 문서로부터 227개의 문서와 dbpedia로부터 61,714 개의 RDF 트리플을 수집하였고, 엔터티 얼라이먼트(혹은 정렬)의 accuracy, recall, 및 f1-score에 대한 비교실험을 수행하였다.

☞ 주제어: 자동 표적 인식, 텍스트-이미지 그래프 변환, 그래프 엔터티 얼라이먼트, 지식 그래프 기반 표적 인식

### ABSTRACT

Automatic Target Recognition (ATR) technology is emerging as a core technology of Future Combat Systems (FCS). Conventional ATR is performed based on IMINT (image information) collected from the SAR sensor, and various image-based deep learning models are used. However, with the development of IT and sensing technology, even though data/information related to ATR is expanding to HUMINT (human information) and SIGINT (signal information), ATR still contains image oriented IMINT data only is being used. In complex and diversified battlefield situations, it is difficult to guarantee high-level ATR accuracy and generalization performance with image data alone. Therefore, we propose a knowledge graph-based ATR method that can utilize image and text data simultaneously in this paper. The main idea of the knowledge graph and deep model-based ATR method is to convert the ATR image and text into graphs according to the characteristics of each data, align it to the knowledge graph, and connect the heterogeneous ATR data through the knowledge graph. In order to convert the ATR image into a graph, an object-tag graph consisting of object tags as nodes is generated from the image by using the pre-trained image object recognition model and the vocabulary of the knowledge graph. On the other hand, the ATR text uses the pre-trained language model, TF-IDF, co-occurrence word graph, and the vocabulary of knowledge graph to generate a word graph composed of nodes with key vocabulary for the ATR. The generated two types of graphs are connected to the knowledge graph using the entity alignment model for improvement of the ATR performance from images and texts. To prove the superiority of the proposed method, 227 documents from web documents and 61,714 RDF triples from dbpedia were collected, and comparison experiments were performed on precision, recall, and f1-score in a perspective of the entity alignment..

☞ keyword: Automatic Target Recognition, Text-image Graph Conversion, Graph Entity Alignment, Knowledge Graph-based Target Recognition

<sup>1</sup> Dept. of Industrial Engineering, Sungkyunkwan University, Suwon, Korea

<sup>2</sup> Agency for Defense Development, Seoul, Korea

\* Corresponding author: myesohn@skku.edu

[Received 30 August 2022, Reviewed 17 September 2022(R2 6 October 2022), Accepted 14 October 2022]

☆ 이 논문은 2022년 정부(방위사업청)의 재원으로 국방과학연구소의 지원을 받아 수행된 연구임(UE201114ED)

## 1. 서 론

미래전투체계(Future Combat Systems, FCS)의 ‘자율성과 지능화’를 달성하기 위한 핵심 요소로 자동 표적 인식(Automatic Target Recognition, ATR)이 부상하고 있다[1]. ATR은 기상조건이나 주야에 관계없이 광범위한 지역의 영상을 고해상도로 수집할 수 있는 Synthetic Aperture Radar (SAR) 센서가 수집한 이민트(IMINT·영상 정보)로부터 전장 상황을 파악하거나 지휘결심을 수행하는 데 필요한 객체 혹은 표적을 자동으로 검출하는 기술을 의미한다[2]. 그러나 수집되는 전장 정보는 이민트에 국한되지 않고 휴민트(HUMINT·인적 정보), 시긴트(SIGINT·신호 정보), 및 트랙보고 등과 같이 다양하기 때문에 이미지 정보만으로 ATR을 수행하는 것은 정보의 손실을 야기할 뿐만 아니라 ATR의 성능 저하에도 부정적인 영향을 미칠 수 있다. 이에 본 논문에서는 이미지와 텍스트 데이터를 동시에 활용할 수 있는 ATR 방법을 제안하고자 한다.

이미지와 텍스트 데이터를 동시에 활용해 ATR을 수행하기 위해서는, 이미지와 텍스트 데이터 간의 관계가 반드시 식별되어야 한다. 일반적으로 이들 간의 관계는 이미지와 텍스트 데이터를 벡터공간에 투사해 의미적으로 유사한 관계를 이미지와 텍스트 데이터를 발견하는 Image-Text Embedding (ITE)을 통해 식별한다[3]. 이때, ITE 연구의 핵심은 대량의 이미지 및 텍스트 데이터와 딥러닝 모델을 활용하여 이들의 의미적인 유사 관계를 표현할 수 있는 Common Semantic Space(CSS)를 발견하는 것이다. 그러나 이미지와 텍스트 데이터만으로는 이들 사이에 존재하는 의미적 불일치성(Semantic discrepancy)으로 인해 일반화된 CSS를 발견하는 것이 어렵다[4]. 이러한 문제를 해결하기 위해, 기존 ITE 연구에서는 이미지와 텍스트 간의 의미적인 유사 관계를 직접적으로 명시된 대규모의 레이블 데이터를 사용하거나 지식그래프와 같이 구조화된 commonsense knowledge를 활용한다[5]. 그러나, 국방 도메인의 경우 다음과 같은 한계로 인해 기존 접근법을 적용하는 것이 어렵다.

국방 도메인은 학습 데이터셋으로 활용할 수 있는 데이터는 제한되며, 특히 이미지와 텍스트 간의 관계를 직접 설명할 수 있는 이미지와 텍스트 데이터의 쌍은 매우 적다.

기존의 ITE 연구는 다양한 유형의 객체나 개념을 포함하는 이미지 및 텍스트 간의 관계를 해석하기 위해 다

양한 딥러닝 기술을 집목한 높은 복잡도의 학습 모델을 활용한다. 그러나, 신속정확한 표적 식별이 목표인 국방 도메인에서 다양한 관계의 학습은 부담만을 야기할 수 있다.

기존 ITE 연구는 예측 성능을 높이기 위해 추상적인 의미를 갖는 이미지 및 텍스트 간의 관계만을 고려한다. 그러나, 국방 분야에서의 정확한 표적 식별을 위해서는 어떠한 객체의 유형이나 모델명보다 더 구체적인 수준에서 다른 객체와 구별되어 유일하게 특정될 수 있는 대상이어야만 한다.

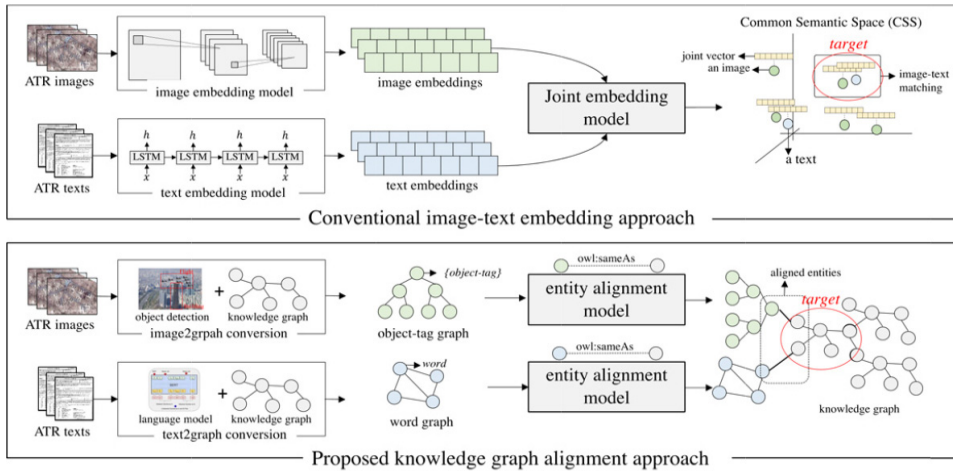
이러한 문제를 해결하기 위해, 본 논문에서 국방 도메인에서 이미지와 텍스트 데이터를 동시에 고려한 ATR을 수행하기 위해, 지식 그래프 기반 이미지와 텍스트를 활용한 ATR 프레임워크를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 연구들을 살펴보고, 3장에서 본 논문에서 제안한 지식 그래프 기반의 ATR방법의 개념을 간략히 설명한다. 4장에서는 제안된 프레임워크의 구조와 각 모듈에 대해 설명한다. 5장에서 실험과 그 결과에 대해 설명하고 6장에서 결론 및 향후 연구에 대해 논의한다.

## 2. 관련 연구

대부분의 ATR 연구는 SAR이나 IR이 관측한 이미지에 대해 이미지 기반의 딥러닝 모델을 활용하여 인식을 수행한다 [2]. 이 연구의 목적은 ATR에 적합한 CNN 및 R-CNN에 기반의 정밀하고 복잡한 딥러닝 모델을 제안하여 보다 정확하고 일반화된 ATR을 수행하는 것이다[6, 7]. 최근에는, 이미지 센싱 기술의 발전으로 인해 수 많은 유형의 이미지 데이터가 생산됨에 따라 복수개의 이미지를 활용하여 ATR을 수행할 수 있는 방법에 관한 연구가 제안되었다[8]. 이미지를 활용한 ATR 연구의 특징은 고해상도의 복잡한 객체들이 포함되어 있는 이미지로부터 신속하고 정확하게 표적만을 식별하는 것이다.

ATR 연구의 상당수는 이미지에 기반하지만 표적과 관련된 데이터의 양과 질이 향상됨에 따라 이미지 이외의 정보를 함께 활용하는 연구도 최근 소수 제안되기도 했다[9]. 이 방법은 ATR보다도 컴퓨터 비전에서 멀티 모달 객체 인식을 위해 제안된 것으로 현재도 활발하게 연구되고 있다[10]. 이 방법의 가장 큰 특징은 이미지와 텍스트같이 상이한 유형의 정보들의 의미 관계를 해석할 수 있는 CSS를 딥러닝 모델을 활용하여 발견하는 것이



(그림 1) 기존의 이미지-텍스트 임베딩 접근법(상단)과 제한된 지식 그래프 얼라이먼트 접근법(하단) 비교  
 (Figure 1) Comparison between conventional image-text embedding (upper) and the proposed knowledge graph alignment (lower) approaches

다. 그러나, 이러한 매니폴드 공간(manifold space)을 발견하는 것은 상당히 많은 규모의 레이블 정보를 요구할 뿐만 아니라 훈련의 복잡도도 매우 높다는 한계가 있다. 이러한 한계를 완화하기 위해 그래프 구조를 활용한 연구들이 제안된 바 있으나[11], 아직은 초기 단계에 머물러 있으며 이질적인 그래프 구조를 해석하는 어려움으로 낮은 일반화 성능을 나타내는 한계가 있다.

### 3. 이미지-텍스트의 의미 관계 식별

일반화된 CSS를 발견하기 위해 수행되는 ITE 절차를 요약하면 그림 1 (상단)과 같다. 이 절차의 가장 큰 특징은 이미지와 텍스트에 적합한 딥러닝 모델을 개별적으로 적용해 최적 임베딩 벡터를 산출한 후, 두개의 임베딩 벡터를 비교가능한 동일한 공간, 즉 CSS로 투영할 수 있는 합동 임베딩 모델(joint embedding model)을 학습하는 것이다. 이 때, 대부분의 합동 임베딩 모델은 일반화된 CSS를 발견하기 위해 이미지와 텍스트 데이터 쌍으로 표현된 의미적 관계 정보를 학습의 레이블 데이터로 요구한다. 그러나 합동 임베딩 모델의 학습에는 상당한 계산 부담이 있으며, 이미지와 텍스트 간의 관계를 직/간접적으로 유추할 수 있는 정보가 매우 희박하고, 발견된 CSS는 특정 이미지와 텍스트 데이터 간의 유사 관계만을 제한적으로 표현하기 때문에 미지의 표적에 대한 ATR에는 한계가 있다.

이러한 한계를 극복하기 위해, 본 논문에서는 지식 그래프 얼라이먼트를 이용한 이미지와 텍스트의 의미 관계(semantic relationship) 식별 방법을 제안한다. 이미지와 텍스트 데이터에 대해 개별적인 임베딩 벡터를 획득하는 대신, 이미지와 텍스트 데이터를 구조화한 그래프로 그래프를 활용하는 이유는 이미지 및 텍스트 데이터로부터 표적과 직접 관련이 있는 부분만이 아니라 ATR의 성능의 높이는 데 기여할 수 있는 표적과 간접적으로 연관되어 있는 정보도 활용하기 위해서이다. 예를 들어, 표적으로 인식해야 할 특정 비행체가 이미지 및 텍스트 데이터에서 관측되었을 때, 해당 비행체의 모델명이나 외관 구조의 정보만으로 관측된 비행체가 물리적으로 유일한 표적임을 판단하기 어렵다. 그러나 비행체와 관련된 엔티티의 정보로 비행 환경, 근접 랜드마크, 관측된 지형정보와 같은 것이 함께 제공되면 유일한 표적임을 인식하는 것이 훨씬 용이할 수 있다. 다시 말해, 표적을 포함하여 표적 관련된 엔티티들은 이미지와 텍스트 간의 관계를 직/간접적으로 유추할 수 있는 공유 정보의 제공을 통해 ATR의 성능 제고에 기여할 수 있다. 본 논문에서 제안한 지식 그래프 얼라이먼트를 이용한 이미지와 텍스트 관계 식별 방법을 도식화하면 그림 1과 같다.

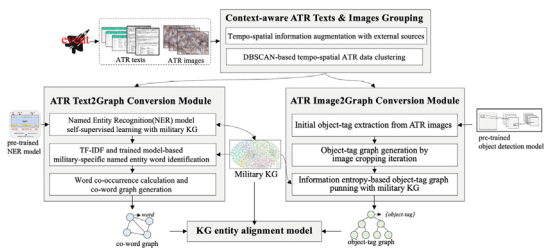
그림 1의 상단에 도식화된 것과 같이 기존의 방법은 이미지와 텍스트 데이터를 개별적인 모델을 활용하여 임베딩 벡터(혹은 공간)로 축약하고 이들 간의 관계를 해석할 수 있는 공간을 다른 모델을 활용하여 학습한다. 이

는 다른 차원으로 투영된 두 공간을 다시 결합하는 것이기 때문에 상당한 계산 복잡도를 요구하며 이 과정에서 큰 정보 손실이 야기된다. 이에 반해, 그림1의 하단에 도식화된 제안 방법은 이미지와 텍스트 데이터를 임베딩 벡터와 같은 공간이 아닌 그래프로의 변환을 선행적으로 수행한 후 변환된 두 그래프를 엔터티 얼라이언트 모델을 통해 결합한다. 이 과정은 그래프를 직접적으로 비교하기 때문에 공간을 결합해야 하는 기존의 방법(상단)에 비해서 계산 복잡도를 낮출 수 있을 뿐만 아니라 정보 손실도 최소화할 수 있다.

제안된 방법에 대한 보다 상세한 내용은 다음과 같다. 이미지 및 텍스트 데이터를 그래프 구조로 변환하기 위해, 데이터 유형에 적합한 사전 학습된 딥러닝 모델과 지식 그래프를 활용한다. 이때, 사전 학습된 딥러닝 모델은 비정형화 된 이미지 및 텍스트 데이터로부터 객체 태그(object tag)나 핵심 단어와 같은 정보를 추출하여 정형화된 구조의 그래프를 생성하는데 필요한 핵심 정보를 해석하기 위해 활용한다. 또한, 지식 그래프의 어휘 정보는 이미지 및 텍스트 데이터로부터 그래프를 생성하는 과정에서 두 그래프 간의 어휘 이질성을 최소화하기 위해 활용한다. 지식 그래프에 기반하여 이미지 및 텍스트로부터 생성된 그래프들은 엔터티 얼라이언트 모델에 입력되어 가장 유사한 의미를 나타내는 지식 그래프의 엔터티와 연결된다. 최종적으로 이미지 및 텍스트 데이터로부터 생성된 그래프를 지식 그래프에 연결하면 표적 인식과 직/간접적으로 관련된 정보가 그래프의 형태로 표현된다.

#### 4. 지식그래프 기반 ATR 프레임워크

본 논문에서 제안하는 이미지와 텍스트 데이터를 활용한 지식그래프 기반 ATR 프레임워크의 구조를 도식화



(그림 2) 이미지와 텍스트 데이터를 이용한 자동 표적 인식의 제안된 프레임워크

(Figure 2) The proposed framework of knowledge graph-based ATR using image and text data source

하면 그림 2와 같다. 제안 프레임워크는 특정 이벤트와 관련되어 입력된 ATR 텍스트 및 이미지를 컨텍스트에 기반하여 그룹핑하는 단계와 이미지와 텍스트 데이터를 각각 그래프로 변환하는 단계로 구성된다.

#### 4.1 컨텍스트 기반 ATR 데이터 그룹핑

특정 이벤트 발생시, 그 이벤트에는 다수의 표적 및 표적 관련 엔터티가 포함되어 있을 수 있다. 이때, 모든 표적 및 표적 관련 엔터티에 대해 ATR를 수행한다면 정보 처리의 부담을 야기할 수 있다. 이러한 부담을 줄이기 위해, 본 논문에서는 특정 이벤트와 관련된 이미지 및 텍스트 데이터를 컨텍스트, 특히 시공간 컨텍스트에 기반하여 그룹핑을 수행한다. 이를 통해, 시공간적으로 관계를 맺기 어려운 표적 및 표적 관련 엔터티간의 불필요한 정보 처리 부담을 줄일 수 있다. 그룹핑을 수행할 ATR 텍스트 및 이미지는 다음과 같이 정의한다.

정의 1.  $i^{th}$  데이터( $d_i$ )  $d_i$ 는 특정 이벤트내 미인식 표적과 관련된 정보를 포함하고 있을 것으로 예상되는  $i$ 번째 이미지 혹은 텍스트 데이터로 다음과 같이 표현된다.

$$d_i = \{c_i, (t_i | p_i)\}, i \leq N \quad (1)$$

이 때,  $c_i$ 는  $d_i$ 가 관측된 미인식 표적에 대한 시공간 컨텍스트 데이터로 미인식 표적의 위치 좌표( $s_i = \{la_i, lo_i\}$ )와 타임스탬프( $t_i$ )로 구성되어 있다( $c_i = \{s_i, t_i\}$ ).  $t_i$ 는 word sequence로 전처리된 미인식 표적에 대한 텍스트 데이터이며,  $p_i$ 는 미인식 표적에 대한 하나의 이미지 데이터이다. 또한,  $d_i$ 는  $t_i$  및  $p_i$ 에서 하나만을 원소로 갖는다( $p_i = \{\emptyset | t_i \in d_i\}$ , or vice versa).

$d_i$ 의 시공간 컨텍스트 데이터  $c_i$ 를 활용하여 다른 ATR 데이터  $d_j$ 와의 시공간 유사성을 비교한다. 그러나,  $c_i$ 는 미인식 표적에 대한 최소한의 시공간 데이터만을 포함하고 있어, 이 데이터만으로는 복잡하고 비선형적이며 non-Euclidian 공간으로 표현되는 ATR 데이터의 시공간 유사성을 식별하기 어렵다. 이러한 문제를 해결하기 위해,  $c_i$ 에 기반하여 시공간 정보를 확장한다.  $c_i$ 의 공간 정보 확장은 위치 좌표( $s_i$ )와 공간 정보에 대한 데이터 베이스(e.g. GeoNames)를 활용해 수행한다. 이를 통해, 미인식 표적이 관측된 지명, 가까운 랜드마크 및 지형지물 정보 등을 획득해  $s_i$ 에 추가하고 시간 정보 확장은 타임스탬프  $t_i$ 와 웹서비스를 활용하여 미인식 표적이 관측된

시점의 기상정보와 같은 정보를 기존  $t_i$ 에 추가한다. 시공간 확장 과정을 통해 미인식 표적과 관련된 시공간 정보가 추가된  $c_i^+$ 를( $|c_i| < |c_i^+|$ ) 획득한다. 시공간 정보가 추가된  $c_i^+$ 를 포함한 데이터는 다음과 같이 표현된다.

정의 2.  $i^{th}$  시공간 확장 데이터( $ad_i$ ,  $ad_i$ 는 시공간 정보가 추가된  $d_i$ 로 다음과 같이 정의된다.

$$ad_i = \{c_i^+, (t_i | p_i)\}, i \leq N \quad (2)$$

이 때,  $c_i^+$ 는 시공간 정보가 보강된 상황데이터이다.

$ad_i$  (for  $\forall i$ )를 이용해 클러스터링을 수행하여 유사 시공간 ( $c_i^+$ )에 대한  $k$ 개의 클러스터를 획득한다. 이 때, 클러스터링은 클러스터의 모양이나 크기에 강건하여 시공간 데이터에 자주 활용되는 DBSCAN을 활용해 수행한다[12]. 또한, 클러스터링에 입력될 시공간 데이터  $c_i^+$ 는 숫자와 명목 변수가 혼합된 데이터이기 때문에 필연적으로 고차원성을 나타낸다. 고차원성을 피하기 위해 모든 시공간 데이터  $c_i^+$  ( $\forall i$ )를 임베딩하고 임베딩 벡터를 활용하여 DBSCAN 기반의 클러스터링을 수행한다. 클러스터링의 수행결과로 ATR 텍스트 및 이미지가 포함된  $J$ 개의 클러스터를 획득한다.

$$C_j \subset \{d_{ijk} | \forall i, k, 2 \leq j \leq J, \sum_j |C_j| = N \quad (3)$$

이 때,  $d_{ijk}$ 는  $j$ 번째 클러스터에  $k$ 번째 인스턴스로 분류된  $i$ 번째 데이터이다.

## 4.2 ATR Text2Graph 변환 모듈

이 모듈의 목적은  $C_j$ 에 포함된 텍스트 데이터 집합  $T_j$  ( $T_j = \{t_{ijk} | t_{ijk} \in C_j, \forall i, k\}$ )를 이용해 이를 대표할 수 있는 그래프를 생성하는 것이다. 이를 요약하면 다음과 같다.

Step 1: Named Entity Recognition(NER) 방법을 이용해  $T_k$ 로부터 실제 표적과 관련된 엔터티 식별

Step 2: 현재 이벤트와 관련된 핵심 Named Entity (NE)를 발견하기 위해 TF-IDF를 수행

Step 3: 핵심 NE를 노드로 간주하는 그래프를 생성하기 위해 핵심 NE의 동시 출현 빈도(co-occurrence)를 활용한 co-word 그래프를 생성

이 과정을 통해 집합  $T_j$ 에 대한 co-word 그래프  $T_j^g$ 를 획득할 수 있다.

### 4.2.1 국방 지식그래프를 활용한 NER 모델의 자기지도 학습

NLP 분야에서 Transformer에 기반한 다수의 NER 모델들이 제안된 바 있다 [13]. 이 때, Named Entity(NE)란 현실 세계에서 고유하게 식별되는 개체 또는 개인을 의미한다. NER 모델의 대부분은 일반적인 도메인이나 말뭉치를 이용해 학습되었기 때문에 국방 분야와 같은 특정 도메인의 데이터에 대해서는 추가적인 학습이 필요하다 [14]. 그러나 NE에 대한 학습 데이터셋을 구축하거나 최신화 하는 작업은 상당한 비용을 요구한다. 따라서, 본 논문에서는 기 구축된 국방 지식 그래프(Military Knowledge Graph, MKG)를 활용하여 기존 NER 모델이 국방에 적합한 모델로 자기지도 학습할 수 있는 방법을 제안한다.

제안 절차는 다음과 같다. 첫 번째, 사전 학습된 NER 모델을 활용하여 MKG에서 NE와 NE가 아닌 어휘를 식별한 후, 이를 원본 또는 실제 값(ground-of-truth) 셋으로 지정한다. 두 번째, 그래프의 구조에 따라 MKG에서 NE와 유사한 그래프 특징을 갖는 후보 NE를 MKG에서 샘플링하고 이를 다시 NER 모델에 입력하여 학습을 수행한다. 세 번째, 기존 NER 모델과 새롭게 학습된 NER 모델에 원본 또는 실제 값 셋을 활용하여 정확도를 평가한다. 새롭게 학습된 NER 모델의 정확도가 기존에 비해 크거나 같다면 두 번째 과정부터 반복한다. 연속적으로  $\rho$ 회 이상 새로 학습된 NER 모델의 성능이 기존보다 좋지 않다면 학습 과정을 종료한다. 최종적으로, 국방 분야에 적합한 Military-specific NER 모델을 획득한다.

### 4.2.2 TF-IDF 기반 핵심NEs 발견

Military-specific NER 모델을 활용하여  $T_j$ 의 텍스트 데이터로부터 식별한 모든 NE가 이벤트와 관련된 핵심 NE가 아닐 수 있다. 이벤트와 관련된 핵심 NE를 선별하기 위해 TF-IDF를 활용한다. Military-specific NER 모델을 활용하여 집합  $T_j$ 에 속한 개별 텍스트 데이터  $t_{ijk}$ 에 대해 식별된 NE의 집합은 다음과 같이 나타낸다.

정의 3. Named Entity 집합 ( $NE_{ijk}$ )  $NE_{ijk}$ 는 텍스트 데이터  $t_{ijk}$ 에 대해 Military-specific NER 모델을 적용하여 식별된 NE의 집합이다.

$$NE_{ijk} = \{ne_{ijk1}, \dots, ne_{ijk2}, \dots | 2 \leq l\} \quad (4)$$

이 때,  $ne_{ijk}$ 는 텍스트 데이터  $t_{ijk}$ 로부터  $l$ 번째 NE로 식별된 단어이다( $ne_{ijk} \in t_{ik}$ ).

모든  $t_{ijk}$ 에 대해 Military-specific NER 모델을 적용하여 복수개의  $NE_{ijk}$ 를 획득한 후 이의 모든 원소  $ne_{ijk}$ 를  $k$  및  $j$ 에 대해 병합하여, 이의 어휘로만 구성된 집합  $\overline{NE}_{ij}$ 를 생성한다.

$$\overline{NE}_{ij} = \{\overline{ne}_t | \forall t, 2 \leq t \leq M\} \quad (5)$$

모든  $NE_{ijk}$ 를 활용하여 어휘집  $\overline{NE}_{ij}$ 의 특정 어휘  $\overline{ne}_t$ 가 집합  $T_j$ 에서 등장할 빈도를 산출한다.

$$tf(\overline{ne}_t) = \sum_{(k,l)} |\{ne_{ijkl} | \overline{ne}_t \equiv ne_{ijkl}\}| \quad (6)$$

이후, 그룹을 망라한 모든 텍스트 데이터  $t_i (\forall i)$ 를 활용하여  $\overline{NE}_{ij}$ 의 특정 어휘  $\overline{ne}_t$ 의 inverse document frequency를 산출한다.

$$idf(\overline{ne}_t) = \log \left( \frac{N}{\sum_i |\{\overline{ne}_t | \overline{ne}_t \in t_i\}|} \right) \quad (7)$$

식 (6) 과 (7)을 활용하여 특정 어휘  $\overline{ne}_t$ 의 TF-IDF 스코어는 다음과 같이 산출한다.

$$tfidf(\overline{ne}_t) = tf(\overline{ne}_t) \times idf(\overline{ne}_t) \quad (8)$$

$tfidf(\overline{ne}_t)$ 의 내림차순 정렬에서 상위  $\alpha\%$ 에 속한  $\overline{ne}_t$  어휘를 제외한 모든 어휘를  $NE_{ijk} (\forall k)$ 에서 제거하여 핵심 NEs의 집합인  $NE_{ijk}^*$ 를 획득한다.

#### 4.2.3 Co-word 그래프 생성

ATR과 관련된 핵심 단어인  $NE_{ijk}^*$ 의 원소를 노드로 하는 co-word 그래프를 생성한다. 그러나 노드만으로 그래프를 생성할 수 없기 때문에 임의의 두 단어가 문서나 문장에 동시에 출현하는 빈도를 산출해 노드 간의 관계를 찾는 word co-occurrence graph 개념을 활용해 co-word 그래프의 관계를 완성한다.

본 논문의 경우, 텍스트 데이터  $t_{ik}$ 에서 임의의 두 어휘  $\overline{ne}_t$ 와  $\overline{ne}_{t'}$ 가 동시에 출현하는 빈도를 다음과 같이 산

출해 co-word 그래프의 관계를 도출한다.

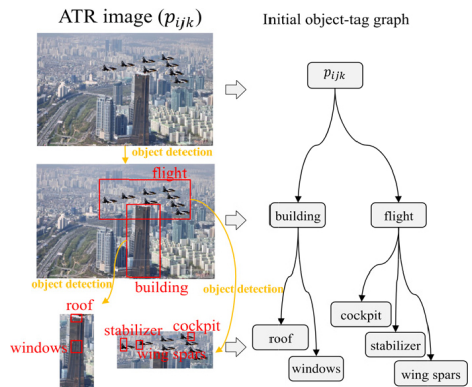
$$co(\overline{ne}_t, \overline{ne}_{t'}) = \sum_{(i,k)} |\{1 | \overline{ne}_t, \overline{ne}_{t'} \subseteq NE_{ijk}^*\}| \quad (9)$$

Co-word 그래프는 노드와 노드 간의 co-occurrence, 즉  $co(\overline{ne}_t, \overline{ne}_{t'})$ 를 원소로 갖는 대칭 행렬  $T_k^g (|T_k^g| = M \times M, T_k^g = T_k^{gT})$ 이다. 이 때, 모든 노드 간의 관계를 맺지 않기 위해서  $co(\overline{ne}_t, \overline{ne}_{t'})$ 의 내림차순 정렬에서 상위  $\beta\%$ 에 속하지 않는 행렬  $T_k^g$ 의 원소 값은 모두 '0'으로 치환한다. 이 때,  $\beta$ 값은 낙차폭이 가장 큰 지점에서 임계 값을 결정하는 elbow method를 활용한다.

#### 4.3 ATR Image2Graph 변환 모듈

이 모듈의 목적은  $C_j$ 에 포함된 이미지 데이터 집합  $P_j (P_j = \{p_{ijk} | p_{ijk} \in C_j, \forall i, k\})$ 로부터 이미지 태그를 획득한 후, 이를 이용해 이미지에 대한 객체-태그 그래프를 생성하는 것이다. 이 때, 객체-태그 그래프의 노드는 인식된 이미지의 객체 태그이며 엣지는 인식된 객체의 이미지 영역을 공유하는 두 이미지 객체-태그 간의 관계를 활용한다. ATR 이미지 데이터로부터 이미지 태그 그래프를 생성하는 과정은 다음과 같다.

첫 번째, 사전 학습된 이미지 객체 인식 모델을 활용하여 이미지  $p_{ijk}$ 로부터 초기 객체-태그 그래프를 생성한다. 이는 이미지  $p_{ijk}$ 에 대한 객체 인식을 수행하고, 객체가 인식된 부분 이미지에 대한 객체 인식의 과정을 반복 수행해 트리 형태의 초기 객체-태그 그래프를 생성한다. 초기 객체-태그 그래프를 생성하는 과정은 그림 3과 같다.



(그림 3) 초기 객체-태그 그래프 생성 과정  
(Figure 3) Process of initial object-tag graph generation

이를 통해 생성된 객체-태그 그래프는 다음과 같이 정의된다.

정의 4. 초기 객체-태그 그래프  $(P_{ijk}^g)$   $P_{ijk}^g$ 는 이미지 데이터  $p_{ijk}$ 에 대해 객체 인식 모델을 반복적으로 적용하여 생성한 그래프이다.

$$P_{ijk}^g = \{V_{ijk}^g, E_{ijk}^g\} \quad (10)$$

이 때,  $V_{ijk}^g$ 는 그래프 노드의 집합으로 이미지 태그  $v_{ijkp}^g$ 를 원소로 가지며( $V_{ijk}^g = \{v_{ijkp}^g | \forall p\}$ ),  $E_{ijk}^g$ 는 그래프 관계의 집합으로 임의의 두 노드 간의 관계를 나타내는  $e_{ijkpq}^g(e_{ijkpq}^g = (v_{ijkp}^g, v_{ijkq}^g))$ 를 원소로 갖는다( $E_{ijk}^g = \{e_{ijkpq}^g | \forall q\}$ ).

두 번째, 지식 그래프와 정보 엔트로피 개념을 활용하여 초기 객체-태그 그래프  $P_{ijk}^g$ 에서 각 노드의 중요도를 산출한다. 초기 객체-태그 그래프는 이미지에서 발견할 수 있는 모든 객체를 반복적으로 식별해 노드로 취했기 때문에, ATR과 관련 없는 노드도 포함될 수 있다. 따라서, 노드의 중요도를 평가하는 과정이 필요하다. 지식 그래프를 활용한 특정 노드  $v_{ijkp}^g$ 의 정보 엔트로피  $H(v_{ijkp}^g)$ 는 다음과 같이 산출한다.

$$v_{ijkp}^g = - \sum_s P(v_{ijkp}^g, s) \log P(v_{ijkp}^g, s) \quad (11)$$

이 때,  $s$ 는 지식 그래프의 샘플링 인덱스이며,  $P(v_{ijkp}^g, s)$ 는  $s$ 번째 샘플링된 지식 그래프에  $v_{ijkp}^g$ 의 어휘가 등장할 확률이다.

마지막으로 식 (11)을 활용하여 모든 산출된 노드의 정보 엔트로피를 산출하고 산출된 엔트로피 값을 코스트로 활용하여 초기 객체-태그 그래프의 가지치기를 수행한다. 가지치기를 수행하는 과정은 다음과 같다. 첫 번째로,  $P_{ijk}^g$ 의 모든 말단 노드를 발견한다. 두 번째, 발견된 말단 노드들을 시작 노드, 최상위의 노드를 도착 노드로 하는 경로를  $P_{ijk}^g$ 로부터 발견한다. 발견된 경로에서 산출된 정보 엔트로피의 변량을 계산한다. 그 다음, 변량이 가장 큰 부분에서 해당 노드보다 깊은 모든 노드를 가지치기 한다. 이 과정을 모든 ATR 이미지 데이터  $p_{ijk}$  ( $\forall i, j$ )에 대해 반복하여 복수개의  $p_{ijk}$ 가 부분 그래프

로 포함된 객체-태그 그래프  $P_k^g$ 를 획득한다.

#### 4.4 ATR KG 엔터티 정렬 모듈

본 모듈에서는 ATR 텍스트 및 이미지로부터 획득된 그래프  $T_k^g$ 와  $P_k^g$ 를 엔터티 정렬(Alignment) 방법을 활용하여 지식 그래프와 연결한다. 엔터티 정렬 방법이란 같은 대상이나 개념을 지칭하나 서로 다른 어휘나 노드로 표현된 두 엔터티 간에 동치관계를 임베딩 및 딥러닝 모델로 예측하는 것을 말한다. 예측을 위해, 특정 노드와 이의 1-hop 이웃 노드를 임베딩하는 Graph Convolutional Network(GCN) 기반의 엔터티 정렬을 수행한다. GCN을 적용한 이유는 이웃 노드의 정보가 충분한  $T_k^g$ 와  $P_k^g$ 의 특성을 반영하기 위해서이다. 두 그래프  $T_k^g$  및  $P_k^g$ 와 지식 그래프 간의 동치 관계를 예측하기 위한 엔터티 정렬은 다음과 같다. 첫째, 동치 관계가 명시되어 있는 모든 엔터티 쌍을 발견한다. 두 번째, 발견한 엔터티 쌍을 학습 데이터의 종속변수로 활용하여  $T_k^g$ ,  $P_k^g$  및 지식 그래프를 GCN 모델에 입력 및 학습한다. 학습된 GCN 모델을 활용하여 그래프  $T_k^g$  및  $P_k^g$ 와 지식 그래프 간의 동치 관계를 예측하고 연결한다.

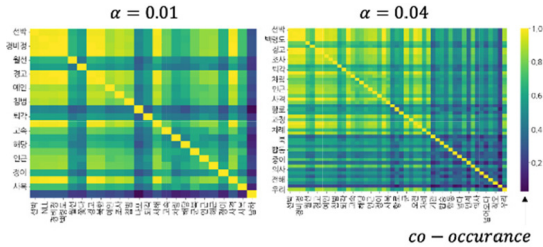
상기 과정을 모든  $k$ 에 대해 반복하여 특정 이벤트에 포함된 미인식 표적에 대한 지식 그래프 기반의 ATR을 수행한다.

### 5. 실험 및 평가

국방 관련 데이터셋에 접근하는 것이 제한되어, 본 논문에서는 국방 ATR 환경과 가장 유사한 "2022년 3월 8일 북한 경비정의 국방한계선(NLL) 침범"과 관련된 기사 및 웹 문서를 ATR 텍스트 데이터로 수집하여 이를 실험 데이터셋으로 활용하였다. 수집된 실험 데이터셋은 평균 단어 개수가 165.69개인 227개의 문서와 지식 그래프의 단위인 61,714개의 RDF(Resource Description Framework)로 구성되어 있다. 또한, 모든 텍스트 데이터는 기본적인 자연어에 대한 전처리를 수행하였다. 모든 실험은 5-cross validation으로 수행된 결과의 평균 값이다.

첫 번째 실험은 하이퍼파라미터 분석이다. 먼저, 하이퍼파라미터  $\alpha$ 는 상위 TF-IDF 값을 갖는 어휘의 선택 비율로 생성되는 co-word 그래프 노드의 수 및 그래프의 크기와 연관이 있다.  $\alpha$  값에 따라 상이하게 나타나는 일부 어휘 간의 동시 출현 빈도 매트릭스는 그림 4와 같다.  $\alpha$

값을 크게 설정할 경우, 의미적으로 중요하지 않은 어휘가 선택될 뿐만 아니라 동시 출현 빈도 값도 낮아서 co-word 그래프의 밀도에 간접적으로 영향을 줄 수 있음을 알 수 있다.

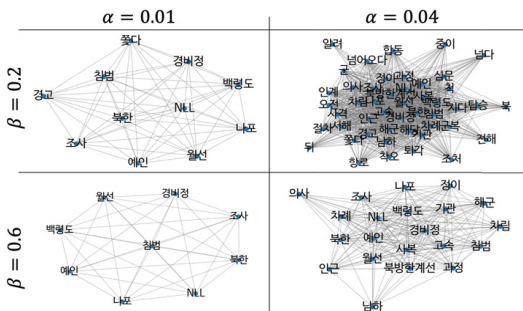


(그림 4)  $\alpha$  값에 따른 동시 출현 빈도 매트릭스

(Figure 4) Co-occurrence frequency matrix by the  $\alpha$

하이퍼파라미터에 따라 변화하는 그래프의 밀도와 형태를 분석하기 위해  $\alpha$  와  $\beta$ 를 동시에 고려해야 한다. 이때,  $\beta$ 는 동시 출현 빈도로부터 그래프에 관계를 부여할 임계 값이다.  $\alpha$ 와  $\beta$ 의 변화에 따라 나타나는 co-word 그래프는 그림 5와 같다.

$\alpha$  값과 반대로  $\beta$  값이 높을수록 보다 단순한 구조의 그래프가 생성된다.  $\beta$  값이 낮으면 대부분의 어휘 간에 관계가 형성되기 때문에 Mesh에 가까운 그래프 구조를 나타낸다. 그러나, 이러한 형태의 그래프는 관계에 기반했을 때 특정 어휘가 다른 어휘와 구분되기 어렵기 때문에 너무 낮은  $\beta$  값은 co-word 그래프의 단어 관계 표현에 악영향을 미친다. 반면,  $\alpha$  값은 높을수록 그래프의 구조 복잡성을 증대되나 보다 많은 어휘 정보를 포함한다. 결과적으로,  $\alpha$  값과  $\beta$  값 모두 다소 높을수록 텍스트의 정보를 변별력 있게 잘 표현할 수 있는 co-word 그래프를 생성할 수 있다.



(그림 5)  $\alpha, \beta$  값에 따라 생성된 co-word 그래프

(Figure 5) Generated co-word graphs by the  $\alpha, \beta$

두 번째 실험은 본 논문에서 제안한 co-word 그래프를 활용한 그래프 기반 엔터티 정렬 방법과 다른 방법과의 비교실험을 수행한다. 엔터티 정렬 방법의 우수성을 평가하기 위해서는 이음동의 관계에 있는 문서 집합 및 지식 그래프 데이터가 필요하다. 이를 위해, 기존 수집된 문서에서 같은 엔터티 명에 대해 설명한 문서를 위키피디아 및 네이버로부터 수집하였다. 또한, 마찬가지로 기존 지식 그래프에서도 같은 엔터티 명에 대한 RDF 집합을 DBpedia와 이 외의 리소스(Yago 혹은 freebase)로부터 수집하였다. 위와 같은 수집 과정으로 85개의 엔터티에 대한 학습 데이터셋을 구축하였다. 제안한 방법과 비교 실험할 방법으로 doc2vec 기반의 문서 임베딩 기법과 비교하였다. Doc2vec의 파라미터로는 임베딩 벡터 크기  $v$  와 윈도우 사이즈  $w$ 를 조정하였으며, 이를 doc2vec  $\{v\}/\{w\}$ 로 표에 나타낸다. Doc2vec는 고차원의 문서를 가장 효과적으로 저차원으로 투영하는 방법 중 하나이기 때문에 본 논문에서 제안한 co-word 그래프와의 우수성을 평가하기 위해 비교 방법으로 선정하였다. 또한, 분류 모델로 서로 접근법이 다른 SVM 및 KNN을 활용하여 분류 모델에 의한 영향을 함께 비교하였다. 실험 결과는 표 1과 같다. 비교 지표로는 정확도(acc), 재현율(recall), 및 f1-score를 활용했다.

(표 1) Doc2vec과의 비교 실험 결과

(Table 1) Comparative experiment result with

Doc2vec			
	acc	recall	f1-score
<b>SVM</b>			
- doc2vec100/4	0.9851	0.5016	0.6375
- doc2vec200/4	0.9825	0.4360	0.5805
- doc2vec100/8	0.9834	0.4732	0.6193
- doc2vec200/8	0.9757	0.3538	0.4949
<b>KNN</b>			
- doc2vec100/4	0.9734	0.02539	0.045
- doc2vec200/4	0.9760	0.1720	0.2824
- doc2vec100/8	0.9747	0.1312	0.2075
- doc2vec200/8	0.9673	0.0847	0.1495
<b>Propose</b>	0.9928	0.8024	0.8876

실험 결과 모든 지표에서 제안한 모델이 가장 좋은 성능을 나타냄을 알 수 있다. 특히, KNN 모델을 활용한 경우 doc2vec의 어떠한 파라미터에서도 낮은 성능을 나타냈다. 그 이유는 임베딩된 벡터의 공간이 KNN 모델이 해석하기 용이한 기하학적 공간이 아니기 때문이기도 하며, 또한 벡터의 크기가 다소 크기 때문에 모델이 일반화에 실패했다. 이는 낮은 재현율 및 f1-score에서 나타난다. 반면, SVM 모델의 경우 doc2vec100/4에서 가장 좋은 성



능을 나타냈으나 마찬가지로 벡터의 사이즈가 커지고 분포가 복잡할 수록( $w$ 가 클수록) 모델의 성능이 저하되었다. 제안된 모델은 높은 재현율 및 f1-score로 보아 일반화에 성공했을 뿐만 아니라 높은 정확도로 다른 방법보다 일반화 및 예측 성능에서의 우수성을 입증했다.

## 6. 결론 및 추후 연구

본 논문에서는 지식 그래프를 활용하여 ATR의 정확도 및 일반화 성능을 향상시키는 지식 그래프 기반 이미지와 텍스트 데이터를 활용한 ATR 프레임워크를 제안한다. 제안된 프레임워크는 ATR 이미지와 텍스트에 대해서 각각 적합한 사전 학습된 모델과 그래프 변환 방법을 통해, 효과적으로 형태가 다른 두 유형의 데이터가 그래프로 얼라이먼트 될 수 있도록 한다. 제안된 프레임워크는 ATR이 이미지에 국한되지 않고 텍스트를 포함한 복수개의 유형의 데이터를 그래프에 기반한 방법을 통해 활용할 수 있도록 하여 정확도와 일반화 성능, 즉 표적에 대한 인식 성능을 직접적으로 향상시키는데 기여한다. 왜냐하면, 저조도 환경이나 운무로 인한 제한된 가시 환경과 같이 이미지 데이터만으로 부족한 표적에 대한 추가 정보를 다양한 센서 및 관측된 데이터가 기록된 텍스트 데이터로부터 획득 및 보완할 수 있기 때문이다.

제안된 프레임워크는 그래프 구조를 활용하여 ATR이 이미지뿐만 아니라 텍스트 데이터도 활용할 수 있도록 기여하나, 다음과 같은 한계가 있다. 첫째, ATR 이미지 및 텍스트로부터 그래프 변환을 수행할 때 지식 그래프의 구조는 활용하지 않고 어휘 정보만을 활용하였다. 둘째, 최종 엔터티 얼라이먼트 과정에서 생성된 텍스트 및 이미지 그래프의 특징을 충분히 고려하지 않은 단일 모델을 활용하였다. 이러한 한계를 극복하기 위해, 추후 연구에서 서로 다른 리소스로부터 발생한 그래프를 일반화된 방법으로 얼라이먼트할 수 있는 그래프 뉴럴 네트워크 모델이나 이의 구조를 지식 그래프와 유사하게 변환하여 얼라이먼트의 성능을 간접적으로 향상시키는 그래프 샘플링 방법에 대해 연구를 수행할 계획이다.

## References

- [1] Matsumurr, J., et al., "Exploring advanced technologies for the future combat systems program," RAND ARROYO CENTER SANTA MONICA CA, 2002. <https://doi.org/10.7249/mr1332>
- [2] Huang, Z., Z. Pan, and B. Lei, "What, where, and how to transfer in SAR target recognition based on deep CNNs," *IEEE Transactions on Geoscience and Remote Sensing*, 58(4), p. 2324-2336, 2019. <https://doi.org/10.1109/tgrs.2019.2947634>
- [3] Mithun, N.C., et al., "Webly supervised joint embedding for cross-modal image-text retrieval," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018. <https://doi.org/10.1145/3240508.3240712>
- [4] Wang, H., et al., "Consensus-aware visual-semantic embedding for image-text matching," in *European Conference on Computer Vision*, Springer, 2020. [https://doi.org/10.1007/978-3-030-58586-0\\_2](https://doi.org/10.1007/978-3-030-58586-0_2)
- [5] Shi, B., et al., "Knowledge Aware Semantic Concept Expansion for Image-Text Matching," in *IJCAI*, 2019. <https://doi.org/10.24963/ijcai.2019/720>
- [6] Kim, S., W.-J. Song, and S.-H. Kim, "Double weight-based SAR and infrared sensor fusion for automatic ground target recognition with deep learning," *Remote Sensing*, 10(1), p. 72, 2018. <https://doi.org/10.3390/rs10010072>
- [7] Xu, C., et al., "An Optimal Faster-RCNN Algorithm for Intelligent Battlefield Target Recognition," in *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, IEEE, 2020. <https://doi.org/10.1109/icaica50127.2020.9181857>
- [8] Zheng, Z., et al., "Dual-path convolutional image-text embeddings with instance loss," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(2), p. 1-23, 2020. <https://doi.org/10.1145/3383184>
- [9] Sakla, W., G. Konjevod, and T.N. Mundhenk, "Deep multi-modal vehicle detection in aerial ISR imagery," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2017. <https://doi.org/10.1109/wacv.2017.107>
- [10] Zhang, D., et al., "Multi-modal graph fusion for named entity recognition with targeted visual guidance," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. <https://doi.org/10.1609/aaai.v35i16.17687>

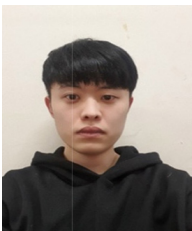
- [11] Lang, C., A. Braun, and A. Valada, "Contrastive object detection using knowledge graph embeddings," Computer Vision and Pattern Recognition, 2021.  
<https://doi.org/10.48550/arXiv.2112.11366>
- [12] Birant, D. and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial - temporal data," Data & knowledge engineering, 60(1), p. 208-221, 2007.  
<https://doi.org/10.1016/j.datak.2006.01.013>
- [13] Yan, H., et al., "TENER: adapting transformer encoder for named entity recognition," Computation and Language, 2019.  
<https://doi.org/10.48550/arXiv.1911.04474>
- [14] Jo, S.-H., et al., "A study on building knowledge base for intelligent battlefield awareness service," Journal of the Korea Society of Computer and Information, 25(4), p. 11-17, 2020.  
<https://doi.org/10.9708/jksci.2020.25.04.011>

## ● 저 자 소 개 ●



### 김 종 모(Jongmo Kim)

2014년 성균관대학교 시스템경영공학과 졸업  
2021년 성균관대학교 산업공학과 졸업 (석박사통합과정)  
2021년 ~ 현재 성균관대학교 산업공학과 박사후연구원  
관심분야 : 그래프 마이닝, 지식그래프, 시맨틱웹, 링크드 오픈 데이터, IOT, 기계학습  
E-mail : dignityc@gmail.com  
[ORCID: 0000-0002-4984-1674]



### 이 정 빈(Jeongbin Lee)

2021년 성균관대학교 시스템경영공학과 졸업  
2022년 ~ 현재 성균관대학교 산업공학과 석사과정 재학 (석박사통합과정)  
E-mail: jim2091@skku.edu



### 전 호 철(Hocheol Jeon)

한양대학교 컴퓨터공학과 공학박사, 국방과학연구소 연구원  
관심분야 : 지휘통제체계, 데이터링크, 상호운용성, 지능형 에이전트, 정보검색, 데이터마이닝  
E-mail: hcjeon71@add.re.kr



### 손 미 애(Mye Sohn)

1985년 성균관대학교 산업공학과(공학사)  
1988년 한국과학기술원 산업공학과(공학석사)  
2002년 한국과학기술원 경영공학과(공학박사)  
2004년 ~ 현재 성균관대학교 산업공학과 교수  
관심분야 : 인공지능/전문가시스템, 지식그래프, 시맨틱웹, 온톨로지, IOT, 기계학습, 추천시스템  
E-mail : myesohn@skku.edu  
[ORCID: 0000-0002-1951-3493]