

BIM 모델 내 공간의 시멘틱 무결성 검증을 위한 그래프 기반 딥러닝 모델 구축에 관한 연구

이원복¹ · 김시현² · 유영수³ · 구본상^{4*}

¹서울과학기술대학교 건설시스템공학과 석사과정 · ²서울과학기술대학교 건설시스템공학과 석사과정 ·
³서울과학기술대학교 건설시스템공학과 박사과정 · ⁴서울과학기술대학교 건설시스템공학과 교수

Development of Graph based Deep Learning methods for Enhancing the Semantic Integrity of Spaces in BIM Models

Lee, Wonbok¹, Kim, Sihyun², Yu, Youngsu³, Koo, Bonsang^{4*}

¹Graduate Student, Department of Civil Engineering, Seoul National University of Science and Technology

²Graduate Student, Department of Civil Engineering, Seoul National University of Science and Technology

³Ph.D. Candidate, Department of Civil Engineering, Seoul National University of Science and Technology

⁴Professor, Department of Civil Engineering, Seoul National University of Science and Technology

Abstract : BIM models allow building spaces to be instantiated and recognized as unique objects independently of model elements. These instantiated spaces provide the required semantics that can be leveraged for building code checking, energy analysis, and evacuation route analysis. However, these spaces or rooms need to be designated manually, which in practice, lead to errors and omissions. Thus, most BIM models today does not guarantee the semantic integrity of space designations, limiting their potential applicability. Recent studies have explored ways to automate space allocation in BIM models using artificial intelligence algorithms, but they are limited in their scope and relatively low classification accuracy. This study explored the use of Graph Convolutional Networks, an algorithm exclusively tailored for graph data structures. The goal was to utilize not only geometry information but also the semantic relational data between spaces and elements in the BIM model. Results of the study confirmed that the accuracy was improved by about 8% compared to algorithms that only used geometric distinctions of the individual spaces.

Keywords : BIM, Space Semantic Integrity, Graph Convolutional Networks, Semantic Relational Information

1. 서론

1.1 연구의 배경 및 목적

국내 건설 산업의 생산성 향상을 위해 전 생애주기에 걸쳐 발생하는 정보들을 효율적으로 관리하는 Building Information Modeling (BIM) 기술의 활용이 원활히 이루어지고 있다. 조달청은 2016년도부터 건설 프로젝트 공공 발주 시 BIM 도입을 의무화하고 있으며, 국토교통부는 건설 산업의 BIM 기술 확산을 위해 2025년까지 모든 건설 산업

에 BIM 기술 적용을 목표로 하고 있다.

이처럼 BIM 도입이 활성화되면서 2D 도면 내 기하학적 폐합면으로만 표현되던 공간은 개별 객체로 인식할 수 있게 되었다. 객체화된 공간의 속성정보는 범규검토, 에너지 분석, 피난 경로 분석 등을 위한 기반 데이터로 사용 가능하기에 BIM의 활용성을 넓힐 수 있는 발판을 마련하였다.

그러나 다수의 참여자에 의해 수작업으로 진행되는 BIM 모델링 과정 특성상 모델 내 기입된 공간정보의 시멘틱 정합성을 보장하지는 못한다. 더불어, 실무에서 활용하는 대부분의 BIM 모델 내에 공간정보가 아예 기입되지 않은 경우도 빈번하여 공간 데이터의 활용이 제한되고 있는 실정이다. 이는 BIM 소프트웨어 사용자가 모델을 검토할 때마다 공간경계를 새로이 정의하고 공간 레이블을 수동으로 수정 및 명기해야 하는 문제를 야기한다(Bloch & Sacks, 2018).

이의 해결을 위해 최근에는 BIM 모델의 공간 정합성 검증

* **Corresponding author:** Koo, Bonsang, Department of Civil Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea

E-mail: bonsang@seoultech.ac.kr

Received January 27, 2022; **revised** March 11, 2022

accepted March 24, 2022

을 위한 연구가 다수 진행되었으며, 대표적으로 추론 규칙 기반 시멘틱 강화 연구(Di Martino et al., 2019)와 기계학습 기반 시멘틱 강화 연구(Bloch & Sacks, 2018; Miller, 2018; Kwon et al., 2019)가 진행된 바 있다. 그러나 이들 연구는 BIM 모델 내 개별 공간의 기하 특성을 일일이 파악하여 규칙으로 구현하는 어려움이 있으며, 기하 특성에만 집중하여 유사한 특성을 가진 공간을 구분하는 데 한계가 있다.

공간은 그 공간을 구성하는 벽, 바닥, 천장 등에 의해 인지 되는 의미론적인 개념으로 인공지능 적용 시 공간을 이루고 있는 관계론적인 의미를 학습 과정에 포함시키는 것이 필요하다(Park & Lee, 2015). 즉, 공간은 공간을 둘러싸는 부재 그리고 인접 공간 간의 관계상 유사성을 띠는 경우가 많다. 예를 들어, 화장실에만 필수적으로 들어가는 부재(세면대, 타일벽 등)가 있기 마련이고, 화장실은 복도와 인접해 있게 된다. 이처럼 공간은 위상적으로 연계된 유사성이 존재하며 이를 그래프나 네트워크로 표현할 수 있다.

이를 반영하기 위해 최근에는 공간 간의 연관 관계를 추가적으로 활용하여 분류 성능을 높이고자 하는 그래프 정보 기반 시멘틱 강화 연구(Wang et al., 2021)가 진행되었다. 해당 연구는 문과 벽을 매개로 추출한 그래프를 학습에 활용하였으나 이에 해당하는 일부 공간(화장실, 거실, 주방 등) 분류에만 적용이 가능하다.

이에 본 연구에서는 공간을 구성하는 개별 BIM 부재와 공간 간 시멘틱 관계로부터 이들 공간의 유사성을 그래프 형태로 표현하고 인공지능으로 학습하여 주어진 제한 공간 분류를 자동화하고자 하였다. 즉, 공간별로 포함하고 있는 부재 집합이 일정 수준 이상 유사할 경우 그래프에서 해당 공간들을 연결함으로써, 연결된 공간의 시멘틱 관계 특성을 인공지능이 학습하도록 하는 것이다. 이를 통해 사무실, 화장실, 복도 등 공간 유형별로 시멘틱 관계 특성이 학습된 인공지능 모델은 분류 대상 공간의 시멘틱 관계정보를 통해 해당 공간의 유형을 추론할 수 있다.

구체적으로, 그래프 기반의 데이터를 분석 및 예측하는 Graph Convolutional Networks (GCN) 알고리즘을 활용하여 공간 분류의 성능을 향상시키고자 하였다. GCN은 합성곱(convolution) 연산의 가중치 공유(weight sharing)와 지역적 특성 학습(local feature learning) 기반으로 노드의 정보와 주변 이웃 노드의 정보를 학습에 활용하여 분류하는 준 지도학습(semi-supervised learning) 알고리즘이다(Kipf & Welling, 2016).

본 연구에서는 공간의 기하정보 뿐 아니라 BIM 모델 내 공간과 부재 간 연결 관계를 GCN 학습과정에 활용하여 향상된 성능의 공간 자동 분류모형을 구축하고자 하였다. 구축된 GCN 기반 모델의 성능을 공간의 기하정보만으로 학습된

기계학습 모델인 Multi-Layer Perceptron (MLP)과 비교하여 공간 분류 시 연결 관계 적용의 효용성을 검증하고자 하였다.

1.2 연구의 범위 및 방법

본 연구는 그래프 기반 딥러닝 알고리즘인 GCN 모델 구축과정에서 시멘틱 관계정보 그래프를 삽입할 경우 향상되는 분류 성능의 정도를 파악하고자 하였다. 이를 위해 12층 규모의 업무시설 유형 BIM 모델 내 9가지의 공간 유형을 대상으로 분석을 진행하였으며, 아래에 제시한 일련의 과정으로 연구를 수행하였다.

1) 공간 레이블 설정 및 학습데이터 구축

연구 대상 BIM 모델에서 총 247개 공간을 추출하고 KBIMS 공간 분류체계를 기반으로 9개의 공간 유형으로 구분하여 학습데이터를 구축하였다.

2) 공간 기하정보 기반 학습모델 구축

베이스라인(baseline) 학습모델 구축을 위해 BIM 모델 내 공간을 대상으로 기하정보(면적, 체적, 길넓이 등)를 추출하였다. 이후 일반 분류기 알고리즘인 MLP를 활용하여 공간의 기하정보만을 이용한 학습모델을 구축하였다.

3) 공간과 부재 간 시멘틱 관계정보 추출

학습에 활용되는 공간과 부재 간 관계정보는 (1) 개별 객체의 경계 상자(bounding-box) 및 (2) 공간경계(space boundary)를 활용하여 추출하였으며, 추출한 관계정보를 GCN 알고리즘 학습에 적합한 유사도 기반 관계정보 데이터 형태로 변환하였다.

4) 시멘틱 관계정보 기반 학습모델 구축

변환한 관계정보 데이터를 활용하여 GCN 알고리즘을 적용한 시멘틱 관계정보 기반 학습모델을 구축하였다.

5) 학습모델 비교 및 검증

베이스라인으로 구축한 공간 기하정보 기반 학습모델(MLP)과 관계정보 기반 학습모델(GCN)의 성능을 비교하기 위해 정확도(accuracy), 정밀도(precision), 재현율(recall) 및 F_1 -score를 평가 지표로 평가 및 검증하였다.

2. 이론적 고찰

2.1 공간 자동 분류 연구

BIM 모델의 품질 신뢰도를 확보하기 위해 모델 내 공간 인식 및 분류를 통한 BIM 데이터 및 IFC¹⁾ 간 매핑 정확

1) IFC (Industry Foundation Classes): 건설정보 호환을 위한 표준 모델로 건설 정보를 중립적이고 개방적인 데이터 포맷으로 제공해 줌으로써, 서로 다른 BIM 소프트웨어 간 객체의 속성 정보 및 형상 정보 등을 교환하는 표준 데이터 모델로 활용됨.

도를 향상시키고, 공간의 정합성을 검증하는 시멘틱 강화 (semantic enrichment) 연구가 다수 수행되었다. 관련 연구는 크게 추론 규칙(inference rule) 기반 연구와 기계학습 (machine learning) 기반 연구로 구분할 수 있다.

추론 규칙 기반 시멘틱 강화 연구는 공간의 기하 특성 및 속성 정보를 기반으로 공간 유형을 유추하는 연역적 방식이다. Di Martino et al. (2019)은 3층 규모의 주거용 아파트 IFC 모델 내 공간 시멘틱 강화를 위해 추론 규칙을 활용하였다. 구체적으로 추론 규칙은 IFC 온톨로지 정보에 기반하여 공간의 기하정보(면적, 체적, 높이 등), 문을 매개로 연결된 공간 간의 연결 관계, 그리고 공간 내 계단의 유무 등을 활용하여 산정하였다. 이후 추론엔진 구축을 통해 특정 공간(거실, 주방, 침실, 화장실 등)의 유형을 예측하고자 하였다.

그러나 추론 규칙 기반 연구는 BIM 모델 내 존재하는 개별 공간의 기하 특성을 일일이 파악하여 규칙으로 구현해야 하는 어려움이 존재한다. 또한, 이를 전문가가 일일이 규명해야 하기 때문에 데이터가 방대해지면 확장성을 보장할 수 없는 한계가 존재한다(Jung et al., 2019).

이에 대한 대안으로 최근에는 기계학습 기반 연구가 수행되었다. Bloch & Sacks (2018)는 추론 규칙 방식과 기계학습 방식의 비교를 위해 각 방식 별로 BIM 모델 내 공간 분류를 진행하였으며, 그 결과 기계학습 방식이 더 우수하다는 결론을 도출하였다. 모든 공간의 특성을 고려하여 규칙을 통해 공간을 구분하는 것은 어려우나, 기계학습을 활용해 공간의 기하 특성(바닥면적, 높이, 개구부의 면적 등)을 학습한 결과 82%의 분류 정확성을 도출하였다. 또한, Proving ground 사는 의료시설을 구성하는 공간 객체에 대해 나이브 베이즈 분류기(naive bayes classifier)를 활용하여 공간 분류를 진행한 바 있다(Miller, 2018). 개별 객체는 북미 분류 체계인 OmniClass Table 13 (spaces by function)에 기반하여 레이블링을 진행하였으며, 공간의 레벨, 형태, 주변 공간의 크기 등을 활용하여 분류를 진행하였다. Kwon et al. (2019)은 주거용 도면으로부터 BIM 데이터를 생성하여 공간속성 정보(상대적 깊이, 통합도, 경로, 면적순위)를 추출하였다. 이후 결정트리(decision tree) 알고리즘을 통해 용도별 규칙을 선정하고, 이를 활용한 추론엔진 구축을 통해 BIM 데이터 내 특정 공간의 용도를 예측하고자 하였다. 해당 연구는 4가지 공간속성 정보를 활용한 규칙을 통해 12가지 유형 공간(주방, 거실, 화장실 등)의 용도 구분을 진행했으며, 약 78%의 정확도가 도출되었다. 해당 연구에서 관련성 높은 속성 정보를 추가로 학습할 경우 더 정확한 구분이 가능해진다고 언급하였다.

그러나 해당 연구들은 공간의 기하 특성에만 집중하여 유사한 특성을 가진 공간을 구분하는 데 한계가 있었다. 기하

정보 학습의 한계를 해결하기 위해 공간의 인접 그래프를 학습에 추가적으로 활용한 그래프 정보 기반 연구가 수행된 바 있다.

대표적으로 Wang et al. (2021)은 아파트 도면 내 문과 벽을 매개로 추출된 그래프를, 그래프 기반 딥러닝 알고리즘인 Graph Neural Network (GNN)에 학습하여 9가지 공간 유형(주방, 화장실, 거실, 침실 등)을 대상으로 공간 분류를 진행하였다. 관계정보의 효용성을 분석하기 위해 다른 위상 및 기하학적 특징은 GNN 모델 학습에서 제외하였다. 학습 결과, 정확도 약 79%의 공간 분류 모델을 구축하였으며, 이는 관계정보를 활용하지 않은 기계학습 모델 대비 약 15% 정도 향상된 수준이었다. 특히, 관계정보 그래프를 추가하였을 때 모든 공간과 연결된 거실(living room)의 정확도가 크게 향상되었다.

그러나 해당 연구는 BIM 모델이 아닌 아파트 도면으로부터 공간 데이터를 취득하였기 때문에, 2차원 정보에 국한되어 있으며 공간의 기하 속성 정보가 학습에 반영되지 않았다. 이에 비해 본 연구에서는 3차원 기하정보 데이터와 유사도 기반 관계정보 그래프를 학습과정에 활용하여, 향상된 분류 성능의 학습모델을 구축하고자 하였다.

2.2 공간 분류 학습 모델

본 연구에서는 BIM 모델 내 공간과 부재 간 관계정보를 학습하기 위해 딥러닝 알고리즘 중 GCN을 활용하였다. 또한, GCN의 성능을 상대 평가하기 위해 기하정보 기반 학습 모델도 학습시켰으며, 이때 기계학습 알고리즘인 MLP를 활용하였다.

2.2.1 Graph Convolutional Networks

기존 딥러닝은 2차원 이미지 정보, 텍스트 등과 같은 Euclidean 정보를 학습에 활용하였으며, 이를 통해 자연어 처리 및 객체 인식 등 많은 연구가 발전하였다. 그러나 포인트 클라우드 데이터 혹은 노드와 엣지로 구성된 그래프와 같은 non-Euclidean 데이터는 기존 딥러닝 기법으로 학습이 되지 않는다. 그래프 데이터는 이미지나 시계열 데이터와 같이 정형적인 형태의 데이터가 아닌 네트워크 형태이므로 인접 노드(node)에 대한 추가 정보를 판독할 수 있어야 한다. 이에 그래프를 정형화하는 graph embedding 연구가 시작되었으며, deepwalk, node2vec, Graph Neural Network (GNN)와 같은 이른바 기하 딥러닝(geometric deep learning) 모델이 다수 등장하였다(Bronstein et al., 2017).

deepwalk와 node2vec 기법은 random walk를 활용하는 기법으로 특정 노드에서 출발하여 일정 범위의 노드를 랜덤하게 선택 후, 해당 노드들을 벡터로 임베딩하여 기존 특성

에 연결 및 분류하는 방식이다(Perozzi et al., 2014). 본 방식은 네트워크 상 이웃 노드 간의 유사 구조를 파악할 수 있다.

GCN은 합성곱(convolution) 연산을 통한 가중치 공유(weight sharing)와 지역적 특성 학습(local feature learning)을 기반으로 노드의 정보와 주변 이웃 노드의 정보를 학습에 활용하는 모델이다(Kipf & Welling, 2016). GCN은 또한 준 지도학습(semi-supervised learning) 방식으로 레이블이 표시된 데이터와 표시되지 않은 데이터를 모두 학습에 활용하여, 레이블이 표시된 데이터의 양이 부족하더라도 높은 학습 정확도를 도출할 수 있다. 즉, GCN은 그래프의 각 노드에 공유된 학습 변수(parameter)를 적용할 수 있으며, 이는 그래프의 크기와 상관없이 해당 노드의 학습 변수만을 사용하게 된다. 따라서 적은 학습 변수만으로도 효과적인 모델을 구축할 수 있으며, 이러한 효율적인 성능으로 주목받은 바 있다.

GCN은 엣지(edge) 간 관계를 나타내는 인접 행렬(adjacency matrix, A)과 노드(node)의 특성을 나타내는 노드 특성 행렬(node feature matrix, H)로 구성된다(Park et al., 2021). 본 연구에서는 인접 행렬에 객체 간 인접관계를, 노드 특성 행렬에는 개별 공간의 기하 특성을 활용하였다. 그래프 내 노드 및 주변 이웃 노드의 학습 변수와 비선형 함수를 통해 신경망(neural network)을 생성한다. 생성한 신경망의 학습 과정을 정리하면 아래의 <Fig. 1, 2>와 같다.

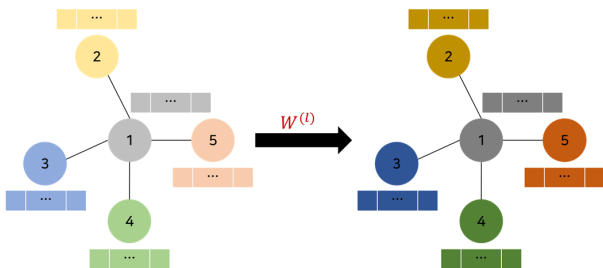


Fig. 1. Process of update node feature in GCN

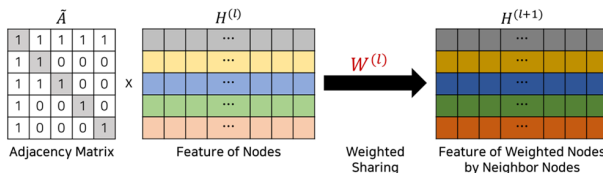


Fig. 2. Learning process of feature matrix

인접행렬과 노드 특성 행렬의 합성곱 연산 및 가중치 공유를 통해 개별 노드와 주변 이웃 노드의 정보를 반영한 새로운 레이어(layer)가 생성되어 학습에 활용된다. 여기서 $w^{(l)}$ 은 l 번째 신경망의 학습 변수를 의미하며, $\tilde{A}=A+I_N$ 는 인접행렬 A와 단위행렬 I_N 을 더한 행렬을 의미한다. 단위행렬 I_N 을 더하는 이유는 인접행렬 A가 주변 이웃 노드와의 연결만 표

현되어 있으므로 학습 과정에서 해당 노드에 대한 특성 정보까지 고려하기 위해서이다(Yang, 2018).

한편, 일반적인 딥러닝 모델은 레이어 수를 증가시켜 skip connection과 같은 방법을 이용하여 모델 성능을 향상시킨다. 이에 반해 GCN의 경우 레이어 수가 증가할수록 임베딩 값이 서로 유사해지면서 정확도가 하락하는 지나친 획일화(over-smoothing) 문제가 발생하므로, 2개의 레이어만을 사용하여 최적의 성능을 도출할 수 있는 2-layer GCN 방식을 학습에 활용하였다. 비선형 함수로 첫 번째 레이어에서는 ReLU와 같은 활성화 함수(activation function)를 이용하였으며, 두 번째 레이어에는 공간의 유형 분류를 위해 softmax를 이용하여 학습을 진행하였다.

2.2.2 Multi-Layer Perceptron

퍼셉트론(perceptron)은 다수의 입력으로부터 하나의 결과로 출력하는 인공 신경망의 기본 유닛(unit)이다. Multi-Layer Perceptron (MLP)은 여러 퍼셉트론의 조합으로 이루어진 network로 입력층(input layer), 출력층(output layer)과 함께 여러 개의 은닉층(hidden layer)으로 구성된 인공 신경망 모형이다. <Fig. 3>에 제시한 MLP의 구조를 보면, 입력층에 해당하는 모든 값들은 모두 은닉층으로 전달된다. 은닉층에서 출력되는 모든 값들은 또한 모두 출력층으로 전달된다. 이때 출력되는 결과값은 활성화 함수(activation function)가 적용된다.

MLP는 비선형 문제를 해결할 수 있으며, 연속적인 값에 대한 예측에 유리하다. 또한 신경망(neural network) 알고리즘의 대표적인 기법이며, 다중 클래스(class) 분류에 주로 활용되므로 기하정보 기반 학습모델에 활용하였다.

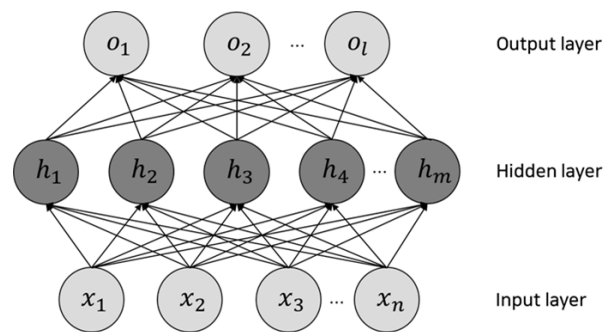


Fig. 3. Multi-layer perceptron structure

3. 연구방법

본 연구에서는 BIM 모델 내 공간의 기하정보 뿐 아니라 공간과 부재 간 연결 관계를 딥러닝 학습과정에 활용하여 공간 분류 성능을 향상시키고자 하였다. 이를 위해 <Fig. 4>에 제시된 절차로 연구를 수행하였다. 우선 해당 BIM 모델

에 공간을 레이블링하고 개별 공간의 기하정보 및 공간과 부재 간 관계정보를 추출하였다. 이후 관계정보 활용의 효율성을 검증하기 위해 기계학습 알고리즘(MLP)은 기하정보만을 학습에 활용하고 딥러닝 알고리즘(GCN)은 기하정보 및 객체 간 관계정보를 학습에 활용하였다. 학습한 모델의 분류 성능 비교 및 분석을 위해 정확도(accuracy, ACC), 정밀도(precision), 재현율(recall) 및 F_1 -score를 활용하였으며, 이를 공간 유형별로 비교 분석하였다.

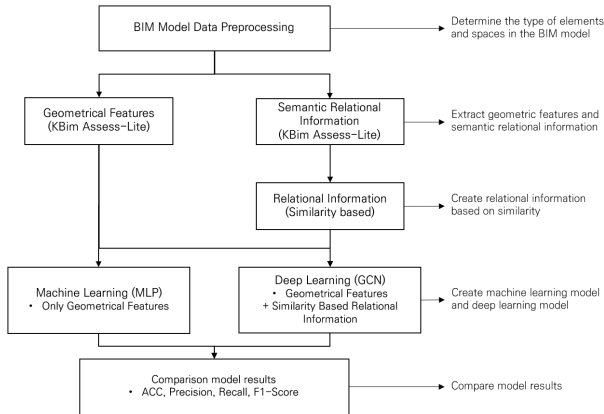


Fig. 4. Process for enhancing space semantic integrity

3.1 연구 활용 BIM 모델 개요

본 연구의 분석대상인 KBIMS 업무시설 모델은 한국 빌딩스마트협회(buildingSMART Korea)에서 제공하는 12층 규모의 LOD (Level of Detail) 300 수준인 실시 설계 단계 표준 IFC 모델로, 9가지 유형의 총 247개 공간과 19가지 유형의 총 11,066개의 부재로 구성되어 있다(Table 1, 2). 연구에서 활용하는 BIM 모델을 구성하는 부재는 beam, covering, wall 등 총 15가지 부재 유형으로 구성되어 있으나, 관계정보 활용 시 특정 부재를 포함한 공간의 분류 성능을 향상시키기 위해 19가지 부재 유형으로 세부 분류를 진행하였다.

Table 1. Status of spaces in BIM model

IFC class	Enumeration type	No. of spaces
IfcSpace	Elevator hall	39
	Elevator vestibule	13
	Hallway	11
	Office room	79
	Plant room	14
	Restroom	20
	Shaft	40
	Stairway	25
	Storage	6
Total		247

Table 2. Status of elements in BIM model

IFC class	Enumeration type	No. of elements
IfcBeam	Beam	158
IfcBuildingElementProxy	Disabled restroom partition	20
	Disabled sign	20
	Elevator door	36
IfcColumn	Column	264
IfcCovering	Covering	65
IfcCurtainWall	Curtain wall	152
IfcDoor	Door etc.	40
	Double door	87
	Single door	88
IfcFlowTerminal	Flow terminal	140
IfcFurnishingElement	Furnishing element	251
IfcMember	Member	6,559
IfcPlate	Plate	1,830
IfcSlab	Slab	303
IfcStair	Stair	23
IfcStairFlight	Stair flight	49
IfcWall	Wall	899
IfcWindow	Window	82
Total		11,066

특정 부재의 세부 분류는 학습모델의 공간 분류성능을 향상시킬 수 있을 것으로 판단되었다. 일례로, <Fig. 5>는 본 연구에서 활용하는 공간 중 승강기홀(elevator hall)과 화장실(restroom)의 예시이다. 해당 공간에서 승강기문(elevator door) 부재는 승강기홀(elevator hall)에만 포함되어 있으며, 양변기·세면대(flow terminal) 부재는 화장실(restroom)에만 포함된 부재이다. 이러한 특정 부재들은 관계정보 활용 시 특정 공간을 분류하는데 유용한 부재로 작용할 수 있으며, 이에 따라 특정 부재들을 세분류하여 분류성능을 높이고자 하였다.

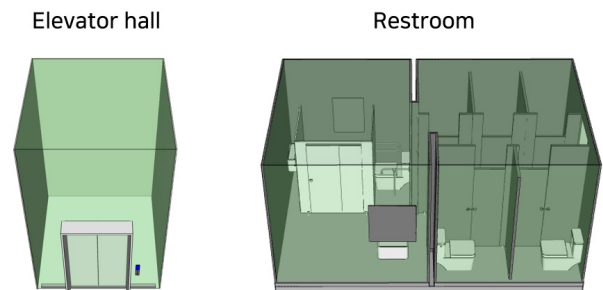


Fig. 5. Example of BIM spaces

3.2 공간 기하 정보 데이터 추출

공간을 분류하기 위한 데이터를 학습하기 위해, 개별 공

간의 기하속성 정보를 특성 변수로 활용하였다. 공간의 기하속성 정보를 추출하기 위해 상용 BIM 소프트웨어인 KBim Assess-Lite 내 탑재된 룰셋(rule set) 파일을 활용하였다. 본 룰셋 파일은 IFC 파일 내 존재하는 공간의 기하속성을 추출하는 기능을 제공하며, 이를 통해 면적(areas), 체적(volumes), 둘레(perimeter), 공간 경계 상자(bounding-box)의 길이·폭·높이(X, Y, Z), 종횡비(aspect ratio), 겉넓이(surface areas), boundary line의 개수, ax1s (area/volume)로 총 10가지를 추출하였다. 해당 특성 변수는 선행연구인 Koo et al. (2018) 및 Bloch & Sacks (2018)에서 활용한 변수를 참조하여 선정하였다.

3.3 시멘틱 관계 정보 데이터 추출

앞선 2.2.1절에 제시한 바와 같이 GCN 학습에는 노드 특성 행렬과 인접 행렬로 두 개의 입력값이 필요하다. 노드 특성 행렬은 공간 자체의 기하 특성을 나타내며 이의 추출은 앞서 3.2절에 제시하였다. 인접 행렬은 개별 공간과 그에 물리적으로 인접해 있는 부재 간의 관계를 나타낸다. 개별 공간은 공간의 용도에 따라 활용하는 부재가 상이하다. 일례로 화장실 공간 내에는 세면대·양변기(flow terminal)를 포함하고 있으며, 승강기실 공간의 경우 승강기문(elevator door)을 포함하고 있다. 기본적으로 공간 내 포함되는 벽(wall), 바닥(slab) 부재 이외에 추가적으로 포함되는 특정 부재와의 관계를 통해 개별 공간의 용도를 예측할 수 있다. 따라서 공간과 부재 간 관계정보를 학습에 활용하고자 하였으며, 이를 통해 특정 부재를 포함한 공간의 분류 정확도 향상을 검증하고자 하였다.

3.3.1 Bounding-box 기반 관계정보 데이터 추출

공간과 부재 간 관계정보 추출을 위해 공간과 부재의 개별 경계 상자(bounding-box)를 활용하였다. 구체적으로, 공

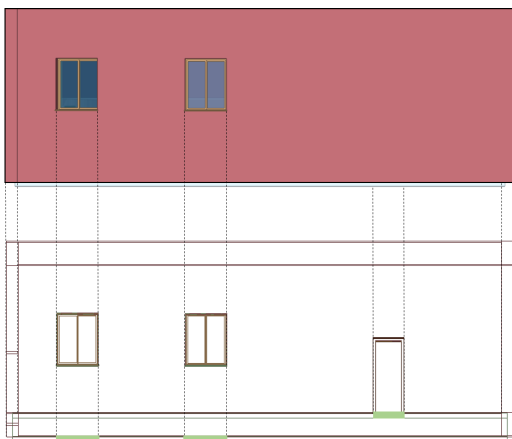


Fig. 6. Extraction of relational information through overlapping bounding-box

간을 구성하는 부재(covering, slab, wall 등)와 이를 포함하는 공간의 경계 상자를 2D 평면에 투영 시 두 투영체 간 중첩이 되는 경우 해당 공간 내에 부재들이 포함되는 것으로 가정하여 물리적 인접관계를 추출하였다(Fig. 6).

3.3.2 공간-부재 인접 그래프 구축

상기 과정을 통해 공간 내 포함 및 인접되어있는 제반 부재를 추출하여 인접 관계 그래프를 구축하였다. 최종적으로 구축된 공간-부재 인접관계 그래프는 11,313개의 노드(node)와 15,199개의 엣지(edge)로 구성되며, 그래프의 형태는 노드 간 방향성이 없는 무방향 그래프(undirected graph)이다. 여기서 노드는 개별 BIM 객체(공간 및 부재)를, 엣지는 물리적 연결 여부를 의미한다.

3.3.3 유사도 기반 관계정보 그래프 구축

앞서 추출한 공간-부재 인접관계 그래프는 객체 간 물리적 인접관계를 활용하여 추출하였다. <Fig. 7>에 제시한 바와 같이 해당 그래프는 공간을 중심으로 공간 내 포함된 부재가 연결되어 있는 형태를 가지고 있다. 그러나 공간 간의 연결 관계가 없으며, 이는 학습과정에 노드 간의 유사성 기반 엣지로 구축된 그래프 형태를 필요로 하는 GCN 알고리즘 적용에 맞지 않는다. 이를 위해 공간의 유사도를 한 공간에 인접한 부재 간의 유사성으로 재정의하고 이를 기준으로 그래프를 재구축하였다.

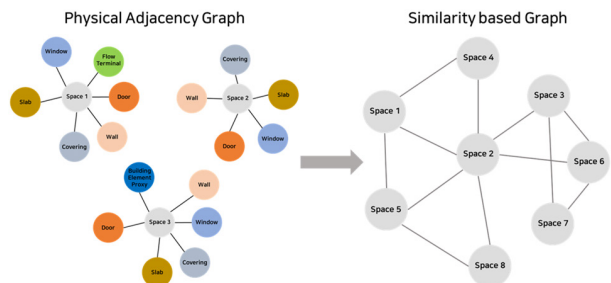


Fig. 7. Extraction of relational information based on similarity

1) 공간 간 유사도 측정

유사도 측정을 위해 유클리디안(euclidean), 자카드(jaccard) 유사도 기법을 조사하였다(Malkauthekar, 2013; Niwattanakul et al., 2013). 이 중 유클리디안 기법은 두 노드 간의 거리를 통해 유사도를 측정하는 방식으로 유클리디안 거리 공식을 이용하여 산정한다. 그러나 본 연구에서 활용하는 관계정보 데이터는 공간과 부재의 좌표를 산정하여 거리를 측정하는 것이 제한적이므로, 해당 기법을 적용하는 것은 적절치 않다.

반면, 자카드 유사도 기법은 집합 간의 교집합 및 합집합을 통해 빈도수를 기반으로 유사도를 측정하는 기법이다. 공간과 이를 구성하는 부재의 연결 관계를 하나의 집합으로

표현하는 것이 가능하므로 해당 기법을 사용하여 유사도를 측정하였다.

구체적으로, 식 (1)에서처럼 개별 공간 내 포함하고 있는 부재를 하나의 집합으로 정의하여 인접하는 공간 내 포함된 부재 총 개수 m 개 ($=A \cup B$) 중 n 개 ($=A \cap B$)가 유사하다면 n/m ($=A \cap B / A \cup B$)의 유사도를 도출하는 방식이다(Fig. 8). 일례로 화장실(restroom) 공간과 승강기 전실(elevator vestibule) 공간의 유사도는 0.25로 계산된다. 화장실 공간은 벽, 바닥, 문, 천장, 칸막이 등 총 7가지의 부재로 구성되어 있으며, 승강기 전실 공간은 벽, 바닥, 승강기문으로 총 3가지 부재로 구성되어있다. 두 공간이 보유한 부재의 총 개수 8개 중 2개의 부재(벽, 바닥)가 유사하므로, 유사도는 $0.25(=2/8)$ 로 도출된다.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

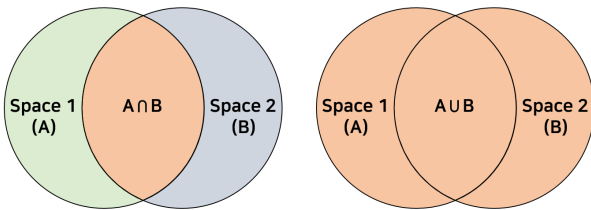


Fig. 8. Jaccard similarity in BIM spaces

2) 유사도 기반 그래프 구축 및 조정

자카드 유사도 기법을 활용하여 유사도를 측정한 결과, 각 유사도 값에 따라 총 61,009개의 엣지로 구성된 그래프가 구축되었다. 그러나 노드 대비 엣지의 개수가 월등히 높아 이에 대한 그래프 조정이 필요하였다.

공간은 이를 구성하는 벽, 바닥 부재를 기본적으로 포함하고 있으며, 용도에 따라 천장, 문, 창문 그리고 양변기, 승강기문 등의 특정 부재가 포함된다.

모든 공간이 기본적으로 벽, 바닥 부재들을 포함하고 있어 유사도가 낮은 공간 간의 엣지를 학습과정에 포함하면 BIM 모델 내 존재하는 모든 공간이 연결된 것으로 나타난다. 이로 인해 학습에 모든 엣지를 활용하였을 때, GCN 학습에 오히려 방해요소로 작용하는 문제가 존재하였다. 이와 관련해서 Li et al. (2019)은 임계점(threshold)을 활용하여 그래프 내 불필요한 엣지를 제거하였다. 해당 연구는 임계점이 0.8에서 1.0 사이의 값으로 적용되었을 때 그래프의 효과적인 성능이 도출된 것을 검증한 바 있다.

따라서 본 연구에서는 선행연구를 참조하여 불필요한 엣지를 효과적으로 제거할 수 있는 임계점을 선정하였다. 최적의 임계점 선정에 위해 0.05 단위로 임계점 수치를 증가시키면서 그래프의 성능을 비교하였다. 그 결과 유사도 임계점이

0.85일 때 4,907개의 엣지가 활용되었으며, 선행연구가 제시한 값의 범위에 가장 근접한 값이 도출되었다. 즉, 불필요한 엣지가 효과적으로 제거된 그래프가 도출되었으며, 이를 모델 학습에 활용하였다(Fig. 9).

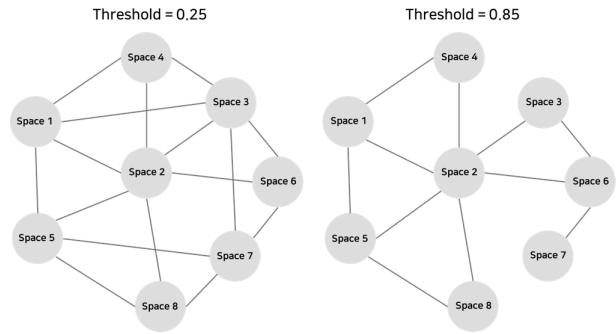


Fig. 9. Similarity based graph according to threshold

3.4 공간 분류 모델 구축 및 검증 방법

앞서 <Fig. 4>에 제시한 대로 학습모델을 이원화하여 구축하였다. MLP는 공간 기하 속성 값들만 활용하여 학습하였으며, GCN은 공간 기하 속성과 자카드 유사도 기법을 활용하여 구축한 유사도 기반 관계정보 그래프를 기반으로 학습하였다. 두 유형의 학습모델은 주어진 BIM 모델 내 9가지 공간 유형에 대해 분류를 진행하였다. 학습모델의 구현은 python의 scikit-learn 기계학습 라이브러리와 pytorch 내 deep graph 라이브러리를 활용하였다.

2가지 학습모델(MLP, GCN)을 학습시키기 위해 3.1절에 소개한 BIM 모델의 공간을 6:4의 비율로 나누어 훈련(train) 및 검증(test) 세트로 활용하였다. 그 결과 <Table 3>에 제시한 바와 같이 148개의 데이터가 모델 훈련, 99개의 데이터가 모델 검증에 활용되었다.

Table 3. Training and test set for each space label

Label	Training set	Test set	Total
Elevator hall	23	16	39
Elevator vestibule	8	5	13
Hallway	7	4	11
Office room	47	32	79
Plant room	8	6	14
Restroom	12	8	20
Shaft	24	16	40
Stairway	15	10	25
Storage	4	2	6
Total	148	99	247

구축한 학습모델의 성능 평가를 위해 정확도(accuracy, ACC), 정밀도(precision), 재현율(recall) 및 F_1 -score를 활용하였다. 정확도는 직관적으로 모델의 성능을 나타낼 수 있

는 평가 지표로 모든 예측 결과 값 중 실제 값을 맞춘 비율을 나타낸다. 정밀도는 양성(positive)으로 예측한 값 중 실제로 양성인 값의 비율이다. 재현율은 실제로 양성인 값 중에서 학습모델이 양성으로 예측한 비율을 나타낸다. F_1 -score는 정밀도와 재현율의 조화평균을 나타내며, 불균형한 데이터의 존재여부를 파악하는 평가 지표로 활용된다.

4. 연구 결과 및 비교

4.1 기하정보 기반 학습모델 검증 결과

기하정보를 활용한 기계 학습모델(MLP) 검증 결과, ACC는 0.86로, F_1 -score는 0.71로 도출되었다. <Table 4>에 MLP 모델 분류 성능을, <Fig. 10>는 precision-recall curve를 제시하였다. 전반적인 분류 성능은 양호한 것으로 확인되었으나, 승강기 전실(elevator vestibule), 설비실(plant room), 창고(storage) 공간의 경우 타 공간에 비해 매우 저조한 성

Table 4. Validation results for MLP

Label	Accuracy	Precision	Recall	F_1 -score
Elevator hall	1.00	1.00	0.76	0.86
Elevator vestibule	0.40	0.40	1.00	0.57
Hallway	0.75	0.75	1.00	0.86
Office room	1.00	1.00	0.94	0.97
Plant room	0.33	0.33	0.67	0.44
Restroom	1.00	1.00	0.89	0.94
Shaft	0.81	0.81	0.76	0.79
Stairway	0.90	0.90	1.00	0.95
Storage	0.00	0.00	0.00	0.00
Average	0.86	0.69	0.78	0.71

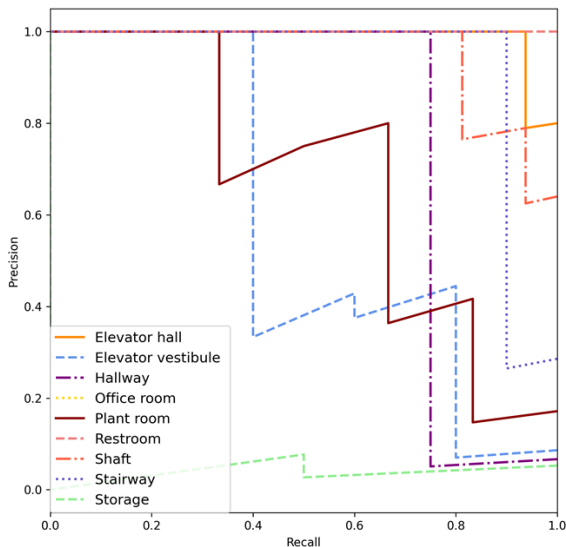


Fig. 10. Precision-recall curve for MLP

능을 보였다. 특히 창고(storage) 공간에 대해 승강기홀(elevator hall)과 계단실(stairway)로 오분류한 결과가 도출되었다. 이는 해당 공간의 학습데이터가 부족하며, 기하 특성이 다른 공간과 명확히 구별되지 않아 낮은 분류 성능을 보인 것으로 확인되었다.

4.2 시멘틱 관계정보 기반 학습모델 검증 결과

시멘틱 관계정보를 활용한 GCN 학습모델의 ACC는 0.94, F_1 -score는 0.91로 도출되었으며, <Table 5>에 공간별 분류 성능 및 <Fig. 11>에 precision-recall curve를 제시하였다. 관계정보를 추가한 GCN 학습모델의 각 공간별 분류 성능은 앞선 기하정보 기반 학습모델보다 우수한 성능을 보였다. 특히 승강기 전실(elevator vestibule)과 창고(storage) 공간의 ACC가 1.00, 0.50으로 상대적으로 향상된 분류 성능을 보였다. 또한 설비실(plant room)과 수직통로(shaft)의 ACC도 0.57, 0.87로 분류 성능이 소폭 향상되었다.

Table 5. Validation results for GCN

Label	Accuracy	Precision	Recall	F_1 -score
Elevator hall	1.00	1.00	0.80	0.89
Elevator vestibule	1.00	1.00	1.00	1.00
Hallway	1.00	1.00	1.00	1.00
Office room	1.00	1.00	0.92	0.96
Plant room	0.57	0.57	1.00	0.73
Restroom	1.00	1.00	1.00	1.00
Shaft	0.87	0.88	1.00	0.93
Stairway	1.00	1.00	1.00	1.00
Storage	0.50	0.50	1.00	0.67
Average	0.94	0.88	0.97	0.91

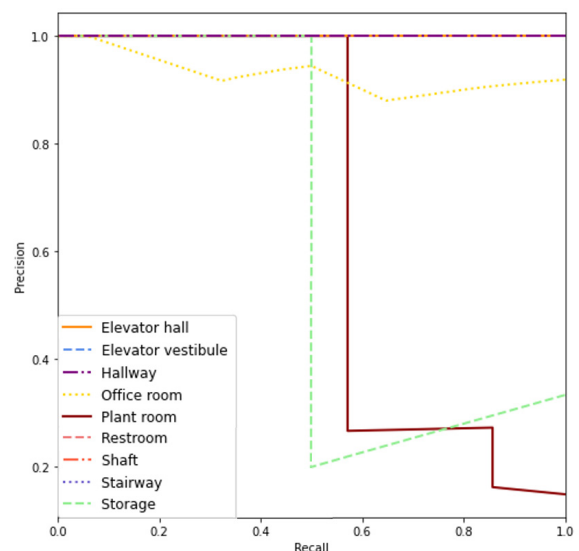


Fig. 11. Precision-recall curve for GCN

4.3 모델 성능 비교 및 오분류 분석

앞서 4.1절에서 제시한 기하정보 기반 MLP 모델의 ACC는 0.86과 F_1 -score는 0.71로 확인되었다. 4.2절에서 제시한 시멘틱 관계정보 기반 GCN 모델의 ACC 및 F_1 -score는 0.94, 0.91로 나타남에 따라 학습과정 내 시멘틱 관계정보를 삽입하였을 때 ACC는 +0.08, F_1 -score는 +0.20의 성능이 향상된 것을 확인하였다.

(Table 6)은 MLP 모델과 GCN 모델의 분류 성능을 비교한 표이다. 제시한 표를 보면 승강기 전실(elevator vestibule), 복도(hallway), 설비실(plant room), 수직통로(shaft), 계단실(stairway), 창고(storage)로 총 6개 공간의 분류 성능이 향상된 것을 볼 수 있다. 특히 공간 내 승강기문(elevator door), 외여닫이문(single door) 부재와 같은 특정 부재를 포함한 승강기 전실과 창고 공간의 분류 성능이 크게 향상되었다. 승강기 전실 공간의 ACC +0.60 (=1.00-0.40) 및 F_1 -score가 +0.43 (=1.00-0.57), 창고 공간의 ACC +0.50 (=0.50-0.00) 및 F_1 -score가 +0.67 (=0.67-0.00)만큼 분류 성능이 향상된 것을 확인할 수 있었다.

그에 반해 다른 공간과 차별되는 부재를 포함하지 않는 설비실과 수직통로 공간의 경우 ACC는 각각 +0.24 (=0.57-0.33), +0.06 (=0.87-0.81), F_1 -score는 +0.29 (=0.73-0.44), +0.14 (=0.93-0.79)로 분류 성능이 소폭 향상된 것을 관측할 수 있다. 특히 수직통로 공간은 형상이 단순하여 기하 특성이 다른 공간과 비교하였을 때 구별되는 특성이 존재하지 않았다. 또한, 수직통로 공간 내 속해있는 그릴창(window)이 모든 수직통로 공간에 포함되어 있지 않아 관계정보의 효과가 저조하였다.

추가적으로 사무실(office room) 공간의 재현율(recall)은 -0.02 (=0.92-0.94), F_1 -score는 -0.01 (=0.96-0.97)만큼 하락하였다. 이는 사무실 공간 내 *IfcFurnishingElement*로 정의되는 부재들이 화장실(restroom) 공간 내에도 존재하는 것에서 기인하였다. 즉, 특정 부재가 두 공간에 동시에 존재함에 따라 GCN 모델이 두 공간을 구분하기 위한 관계정보 상 명확한 차이점을 탐색하지 못하였고, 이로 인해 일부 사무실 공간을 화장실로 오분류한 것으로 보인다.

요약컨대, 특정 공간에 속해 있는 부재들과의 관계를 그래프 형태로 구축한 후 학습시킴으로써 유의미한 분류 성능의 개선이 있다는 것을 확인할 수 있다. 특히 GCN 학습 과정 내 시멘틱 관계정보 활용은 공간 내 포함된 부재가 다른 공간 내 부재와 차별될수록 분류 성능 향상 효과가 높으며, 포함된 부재가 다른 공간과 유사하다면 성능 향상에 한계점이 존재한다는 것을 확인할 수 있다.

Table 6. Delta values between MLP and GCN metrics (MLP-GCN)

Label	Accuracy	Precision	Recall	F_1 -score
Elevator hall	+0.00	+1.00	+0.04	+0.03
Elevator vestibule	+0.60	+0.60	+0.00	+0.43
Hallway	+0.25	+0.25	+0.00	+0.14
Office room	+0.00	+0.00	-0.02	-0.01
Plant room	+0.24	+0.24	+0.33	+0.29
Restroom	+0.00	+0.00	+0.11	+0.06
Shaft	+0.06	+0.07	+0.24	+0.14
Stairway	+0.10	+0.10	+0.00	+0.05
Storage	+0.50	+0.50	+1.00	+0.67
Average	+0.08	+0.19	+0.19	+0.20

5. 결론

본 연구는 BIM 모델의 공간 시멘틱 무결성 검증을 위해 학습모델 구축 시 공간의 기하정보와 물리적 인접 관계정보를 함께 활용하여 높은 분류 정확도를 달성하고자 하였다. 특히 GCN 모델은 단순히 공간의 기하정보를 학습한 MLP 모델 대비 ACC가 약 8% 향상되었으며, 공간의 기하 속성만으로 구분이 어려운 공간들을 적절히 분류하는 것으로 확인되었다. 또한, 기하적 형상이 다른 공간과 크게 차별성이 없었음에도 불구하고, 관계정보 내 차별되는 특정 부재를 포함하고 있어 공간 분류 정확도가 향상되었다.

그러나, 설비실(plant room)과 창고(storage)의 경우 관계정보 상 특징점을 적절히 탐색하지 못하여 정확도의 향상이 크지 못했다. 즉, 학습데이터가 부족하여 해당 공간 특성이 학습과정에 반영되지 않았으며, 이는 해당 공간의 한계점으로 작용하였다. 이를 위해 해당 특성을 모델이 학습할 수 있도록 추가 데이터 수집이 필요하다.

공간 및 부재 간의 관계를 그래프화하여 학습할 때 가장 큰 문제점 중 하나는 단일 BIM 모델을 사용할 수밖에 없다는 것이다. 즉, 각 BIM 모델마다 자체적인 그래프는 생성되나, 여러 BIM 모델을 하나의 그래프로는 표현할 수 없다. 또한, 학습에 용도가 서로 일치하지 않는 BIM 모델을 사용할 경우 공간 유형과 배치 등이 상이하여 학습데이터로 활용하는 데 제약이 존재한다. 그러므로 여러 BIM 모델을 토대로 데이터량을 확장할 수 없는 중차대한 한계가 존재하였다. 이는 단순 물리적 인접관계를 활용하여 그래프를 구축할 경우 생기는 문제이다. 그러나 본 연구에서 제시한 유사도 기반 접근법은 기존 물리적 인접관계에서 추출된 유사도를 활용하여 공간 간의 연결 관계가 존재하는 새로운 그래프를 생성하고 GCN 학습에 활용하기 때문에 상기 한계점에 구애받지 않는다. 이러한 측면에서 데이터를 축적하고 이를 활용

할 수 있는 기반을 마련했다는 면에서 연구의 의미가 있다. 즉, 유사도를 기반으로 그래프를 생성하기 때문에 단일 BIM 모델에 국한되지 않으며, 다수의 BIM 모델을 학습에 활용할 수 있는 이점이 존재한다. 본 연구에서는 실무에서 주로 활용되는 업무시설 용도의 단일 BIM 모델에 집중하여 연구를 진행하였다. 향후에는 추가 BIM 모델 수집을 통해, 다양한 용도의 BIM 모델 내 공간과 부재 간의 관계에 대해 면밀히 분석하여 모델의 성능을 향상시킬 계획이다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.
(No.2020R1A2C1100741).

References

- Bloch, T., and Sacks, R. (2018). "Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models." *Automation in Construction*, 91, pp. 256-272.
- Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017). "Geometric deep learning: going beyond euclidean data." *IEEE Signal Processing Magazine*, 34(4), pp. 18-42.
- Di Martino et al. (2019). "A semantic and rule based technique and inference engine for discovering real estate units in building information models." *In 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering, AIKE*, pp. 81-88.
- Jung, R.K., Koo, B.S., and Yu, Y.S. (2019). "Using deep learning for automated classification of wall subtypes for semantic integrity checking of building information models." *Journal of KIBIM*, 9(4), pp. 31-40.
- Kipf, T.N., and Welling, M. (2016). "Semi-supervised classification with graph convolutional networks." *Conference paper of International Conference on Learning Representations 2017*.
- Koo, B.S., Yu, Y.S., and Jung, R.K. (2018). "Machine learning based approach to building element classification for semantic integrity checking of building information models." *Korean Journal of Computational Design and Engineering*, 23(4), pp. 373-383.
- Kwon, O.C., and Cho, J.W. (2019). "Space Usage Knowledge Extraction from BIM Data by Decision Tree and Expert System." *Korean Journal of Computational Design and Engineering*, pp. 126-134.
- Li, F., Zhu, Z., Zhang, X., Cheng, J., and Zhao, Y. (2019). "Diffusion induced graph representation learning." *Neurocomputing*, 360, pp. 220-229.
- Malkauthekar, M.D. (2013). "Analysis of Euclidean distance and Manhattan distance measure in Face recognition." *In Third International Conference on Computational Intelligence and Information Technology*, pp. pp. 503-507.
- Miller, N. (2018). "AEC Tech 2018 - How to Train your Algorithm." REVIT.NEWS, <<https://www.revit.news/2018/aec-tech-2018-how-to-train-your-algorithm-2/>> (Dec. 15, 2021)
- Niwattanakul, S., Singthongchai, J., Naenudorn, E., and Wanapu, S. (2013). "Using of Jaccard coefficient for keywords similarity." *In Proceedings of the international multiconference of engineers and computer scientists*, 1(6), pp. 380-384.
- Park, J.H., Kim, D.G., Jeong, H.G., and Han, C.E. (2021). "Alzheimer's disease classification using graph convolutional networks." *The Institute of Electronics and Information Engineers*, pp. 2040-2046.
- Park, S.K., and Lee, J.K. (2015). "A Study on the Technological Connections between BIM (Building Information Modeling) and Interior Architecture Design -Focusing on the Applications of Spatial Object and its Properties." *Journal of Korea Design Knowledge*, 34, pp. 35-44.
- Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). "Deepwalk: Online learning of social representations." *In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701-710.
- Wang, Z., Sacks, R., and Yeung, T. (2021). "Exploring graph neural networks for semantic enrichment: Room type classification." *Automation in Construction*, 104039.
- Yang, H.C. (2018). "High Resolution 3D Mesh Object Reconstruction Using Graphical Convolutional Neural Networks." Master's Dissertation, Seoul National University.

요약 : BIM의 도입에 따라 공간이 개별 객체로 인식되면서 객체화된 공간의 속성정보는 범규검토, 에너지 분석, 피난 경로 분석 등을 위한 기반 데이터로 사용 가능하기에 BIM의 활용성을 넓힐 수 있는 발판을 마련하였다. 그러나 BIM 모델 내 개별 공간 속성의 오기 입이나 누락이 없는 시멘틱 무결성(semantic integrity)이 보장되어야 하는데, 다수의 참여자에 의한 수작업으로 진행되는 BIM 모델링 과정 특성 상 설계 오류가 빈번히 발생한다는 문제점이 존재한다. 이를 해결하기 위해 BIM 모델의 공간 정합성 검증을 위한 연구가 다수 진행되었으나, 적용 범위가 한정적이거나 분류 정확도가 낮은 한계점이 존재하였다. 본 연구에서는 공간의 기하정보 뿐 아니라 BIM 모델 내 공간과 부재 간 연결 관계를 Graph Convolutional Networks (GCN) 학습과정에 활용하여 향상된 성능의 공간 자동 분류모델을 구축하고자 하였다. 구축된 GCN 기반 모델의 성능을 공간의 기하정보만으로 학습된 기계학습 모델인 Multi-Layer Perceptron (MLP)과 비교하여 공간 분류 시 연결 관계 적용의 효용성을 검증하고자 하였다. 이를 통해 관계정보 활용 시 약 8% 내외 수준으로 공간 분류 성능이 향상되는 것으로 확인되었다.

키워드 : BIM, 공간 시멘틱 무결성, GCN, 시멘틱 관계정보
