

Construction Action Detector (v0.8)

미시적 개별 건축행위 발생 지역 및 주변지역 특성에 대한 탐색

Chang-ho, Jang^a

^aUrban Engine Team Leader, Urban Engine Project Founder.

Engine Info

개별 건축행위 및 주변 지역의 특성에 대한 시계열적 추적 및 유형화

Keywords:

Urban Planning
Urban Development
Micro Development
Building Construction
Deep Clustering AI Engine

ABSTRACT

개별 건축 행위에 대한 탐색을 주제로 하는 “Construction Action Detector”의 개발 목적은 개별 건축 인허가 행위와 지역 특성을 체계적으로 분석하여 도시 개발의 미시적 단위에서 발생하는 건축 행위를 이해하는 것이다. 다양한 공공 데이터를 통합하여 건축 행위가 발생한 지역의 미시적인 특성을 파악하였다. Urban Engine의 Core Technology인 Deep Clustering AI Engine을 적용하고 DEC 모델을 활용하여 지역별 건축 행위의 특징을 도출하였다. 또한, 시각화 대시보드를 통해 엔진의 시각화 결과를 직관적으로 제공하여 시군구의 미시적 개발 특성을 쉽게 이해할 수 있게 하였다. 향후 본 엔진에서 교통 등과 관련된 추가 변수를 포함하고 모델 개선과 검증을 지속하여 엔진의 활용성을 높일 예정이다. 본 엔진은 도시 개발 전략 수립에 중요한 기초 자료로 활용될 것이며, 지속적인 업데이트와 고도화를 통해 더욱 효과적인 도시 계획 지원 도구로 발전할 것으로 기대된다.

1. Background

도시개발의 관점에서 개별 건축 행위는 가장 기초적이고 미시적인 도시 개발의 기본 단위라 일컬을 수 있다. 이와 같은 개별 건축행위들이 모여, 지역/지구 개발이 되거나 신도시 개발과 같은 대규모의 택지 및 도시개발행위로 나타나게 된다. 이처럼 개발사업의 규모와 관계 없이 개별 건축행위는 도시개발행위와 불가분의 관계에 있다고 말할 수 있으며, 미시적인 건축 행위에 대해서 이해하는 것은 보다 복잡하고 거시적인 도시 개발에 대해 이해할 뿐만 아니라, 도시개발 사업성 평가, 효과성 입증, 부지 선정 등 여러 관련된 문제를 해결하기 위해 효과적으로 활용될 수 있다. 특히, 개발사업지구와 같은 거시적인 지역 및 지구에 대한 개발에 앞서서 인근 지역 혹은 유사한 특성을 지니는 지역에서 발생한 미시적인 개별 건축 개발 행위에 대한 분석을 바탕으로 개발사업에 대한 종합적인 판단까지도 수행할 수 있다.

또한, 도시계획적 관점에서도 미시적인 건축 행위의 특성을

파악하는 것은 큰 의미가 있다. 우선, 도시의 개발 양상과 패턴에 대한 파악을 수행할 수 있다. 보다 구체적으로 살펴보면, 도시의 개발이 증가/감소하는지 파악하는 것을 통해 도시가 성장세를 파악하거나 도시 활력의 감소를 직접적으로 확인할 수 있다. 도시의 인프라 대비 지나친 난개발이 발생하는지를 확인하고 도시 스프롤(Urban Sprawl)에 대해 대비하는데 활용할 수도 있다. 다음으로, 도시에서 더 이상의 신규개발 없이 정체/노후화 되는 지역들을 탐색하고 해당 지역들을 계획적으로 관리하기 위한 정비계획 등을 수립하는데 도움을 줄 수 있다. 각각의 지자체에서 건축행위가 다수 발생하는 지역의 특성을 파악하고, 쇠퇴하는 지역에서의 개선 방향성을 찾는데 충분히 이용할 수 있다.

이처럼 개별 건축 행위에 대해 탐색하는 것은 여러가지 장점을 가지고 있음을 확인하였다. Urban Engine에서도 Urban Planning Engine의 주요 개발 영역 중 하나인 도시의 개발 잠재력을 평가하는데 있어서도 주요한 기초자료로 활용할 수 있다는 점에서 중요한 의미를 가지고 있다.

2. Purpose & Concept of Engine

Construction Action Detector은 미시적인 개별 건축 개발 행위를 탐색하기 위해, 다음을 실현하고자 하는 목적이 있다.

〈Engine Purpose〉

“개별 건축 인허가 행위의 추세와 개발이 발생한 주변 지역의 특성을 이용자들이 효과적으로 탐색할 수 있게 하여, 미시적인 개별 건축 개발 행위의 도시별/지역별 특징을 시계열적으로 파악”

Construction Action Detector의 개발 Concept는 다음과 같다. 특히, 기 개발하는 엔진은 후속 개발 엔진인 도시개발 구역의 잠재력을 평가하는데 활용되는 기초 엔진 성격을 가지고 있다.

〈Engine Concept〉

- 본 엔진은 분석을 목표로 하기 보다는, 향후 도시개발 관련 분석 엔진에서 활용될 수 있는 데이터 수집 및 아카이빙, 현상에 대한 탐색에 초점을 맞춤
- 건축물 대장 인허가 데이터를 기반으로 시군구별 개별 건축행위에 대한 시계열적 탐색 및 공간분포 식별
- 개별 건축행위가 일어난 주변 지역의 특성을 데이터화하고, 지역의 전체적 특성을 진단
- 내부 Core AI Engine인 Deep Clustering Engine을 적용하여, 개별 건축 행위 발생 유형을 도출
- 도시개발 잠재력을 측정하는 AI Engine에서 활용될 수 있는 기초 자료로 활용
- 추후 서비스될 맞춤형 소지역 Cell 단위의 개발 특성 분석 서비스와 연계될 수 있도록 개발 수행

3. Literature Review

Construction Action Detector는 개별 건축행위에 대한 데이터 수집 및 주변 지역의 특성을 탐색하는 것이 주 목적이다. 따라서, 문헌 검토에 대한 필요성은 상대적으로 적으므로, 본 엔진에서는 아직 별다른 문헌 검토를 수행하지 않았다. (추후 추가될 수 있음). 엔진에서 적용된 Deep Clustering Engine에 대한 검토는 Methodology에 별도로 기술하였다.

4. Methodology

1) Urban Engine의 시공간적 범위

Construction Action Detector에서 적용되는 시공간적 범위는 다음과 같다.

- 시간적 범위 : 2018 ~ 2023⁺
(현재 시점 기준, 매년 데이터 업데이트)
- 공간적 범위 : 전국 시군구 및 50M CELL 단위

2) Data

Construction Action Detector의 구현을 위해서, 건축물 대장, 인구, 종사자, 전자지도, 인허가 데이터 등 다양한 소스의 공공데이터를 원천으로 활용하였다. 기본적으로, 개별 건축행위의 위치, 인허가 시점 등을 구체적으로 파악하기 위하여 세움터 건축데이터 민간개방 시스템 (EAIS)에서 제공하고 있는 건축인허가, 주택인허가 대용량 개방 데이터를 활용하였다. 본 엔진에서는 개별 건축 행위가 발생한 지역의 특성을 규명하고자 하는 목적이 있으므로, 건축물이 준공된 이후의 건축물 대장을 활용하기보다는 실제 건축 행위와 관련된 의사결정이 이루어진 시점인 인허가 시점을 활용하기 위하여 건축인허가 및 주택인허가 데이터를 활용하였다.

개별 건축행위가 발생한 주변 지역의 특성을 규명하기 위해 다양한 데이터가 활용되었다. 통계지리정보시스템의 인구, 종사자, 사업체 데이터를 활용하여 인구 및 고용과 관련된 기초 정보를 포함하였다. 또한, 높은 영향도가 있을 것으로 예상되는 공시지가 데이터가 활용되었으며, 이외에도 지목 정보, 건축물 속성 정보, 도로 정보, 시설정보, 지방행정 인허가 데이터 기반의 다양한 파생 데이터가 개발되어 개별 건축행위가 발생한 지역의 다이나믹한 특성을 파악할 수 있도록 하였다. 구체적으로 엔진 개발에 활용된 기초 데이터는 〈Table 1〉과 같다.

〈Table 1〉에 포함된 데이터는 현재까지 Construction Action Detector 및 개발중인 타 분석 엔진에 활용된 기초 데이터가 모두 포함되어 있다. 개별 건축 개발 행위에 주요한 영향을 끼치는 교통 관련 데이터 및 변수들은 아직 수집 및 개발되어 있지 않으며, 해당 영역의 경우 타 엔진을 개발함과 동시에 기 엔진의 지속적인 업데이트를 통해, 추가될 예정이다.

[Table 1] Engine 활용 기초 데이터

데이터 명	설명	출처
건축물 대장 데이터	건축인허가, 주택인허가 대용량 개방 데이터	EAIS (세움터)
인구 데이터	100m 국가격자 단위 인구 분포 데이터	SGIS (통계지리정보시스템)
종사자 데이터	100m 국가격자 단위 종사자 수, 사업체 수 분포 데이터	SGIS (통계지리정보시스템)
공시지가 데이터	필지단위 개별공시지가 데이터 (대지)	국가중점데이터
지목(필지) 데이터	지목별 공간 형상 (필지)	국가중점데이터
건축물 정보 데이터	도로명주소 전자지도 내 건축물 속성 정보 데이터	주소기반산업지원서비스
도로 정보 데이터	도로명주소 전자지도 내 도로 속성 정보 데이터	주소기반산업지원서비스
기타 시설정보 데이터	도로명주소 기타자료(공원, 하천 등) 및 사물주소 시설 도형 정보 데이터	주소기반산업지원서비스
인하가 데이터	업종별 인허가 데이터 (좌표기반 데이터)	지방행정인허가데이터개방

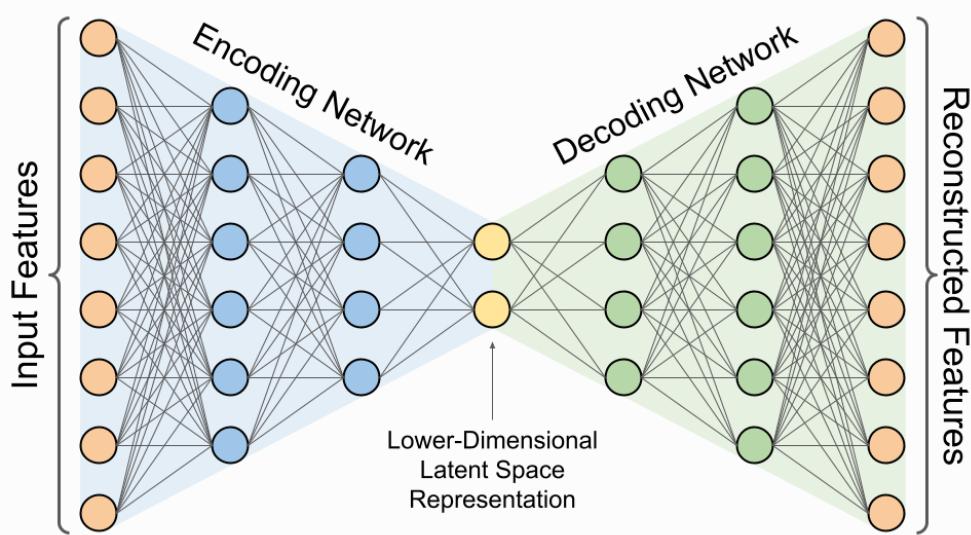
3) Deep Clustering AI Engine

Construction Action Detector에서는 Urban Engine이 보유한 Core 엔진인 Deep Clustering AI Engine을 사용하여, 개별 건축 행위가 발생한 지역에 대한 유형화를 수행하였다. 본 문서에서는 해당 엔진에 대해서 큰 틀에 대해서 서술하였으며, 구체적인 사항들은 추후 공개될 별도의 기술서를 통하여 확인할 수 있다.

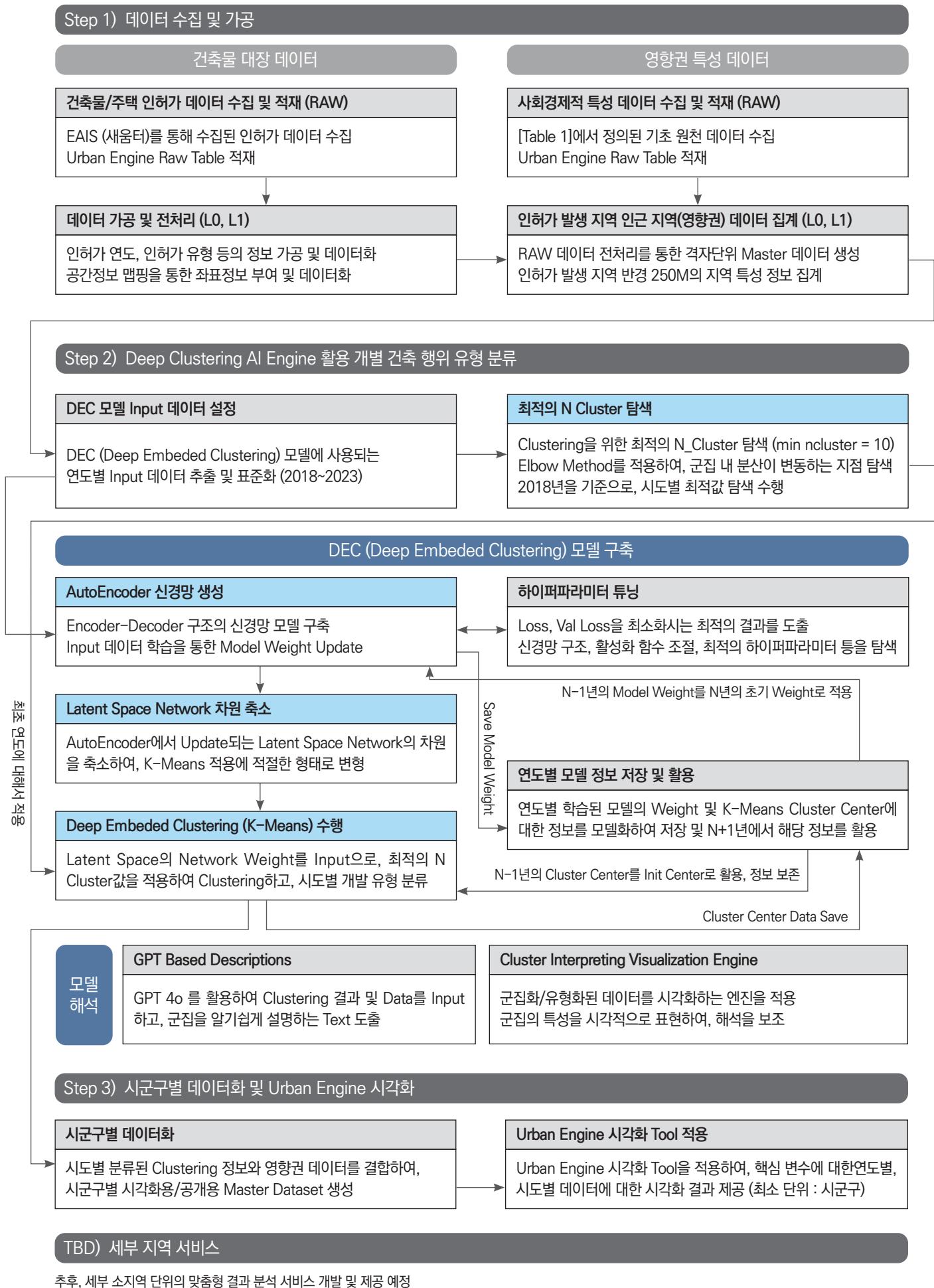
본 엔진에서는 Deep Clustering AI Engine에서 DEC(Deep Embedded Clustering) Model을 통해 데이터의 군집화를 수행한다. 이는 비지도 학습(Unsupervised Learning)의 개념으로, 별도의 레이블링, 속성부여 등의 작업 없이 데이터만을 이용하여 결과를 도출시킬 수 있다. Construction Action Detector에서 DEC모델을 활용한 가장 큰 이유는 군집의 결

과가 시계열적인 안정성을 가질 수 있기 때문이다. 일반적인 통계 기반의 군집분석 방법론인 K-Means, Agglomerative Clustering 등은 N년 대비 N+1, N+2, ... 연도의 군집의 연속성을 보장하기 어렵다는 단점이 있다. 매 연도에 따라서 군집의 크기, 군집의 중심 등이 변화될 수 있고, 특히 시행마다 Random한 특성을 갖는 Clustering Model에서는 이러한 시계열적 안정성이 더욱 줄어들 수 있다. DEC 모델에서는 이러한 단점을 극복하기 위하여, N년의 군집 결과를 도출하는 모델에서 N-1년의 모델의 가중치들을 활용하는 과정을 거친다. DEC의 AutoEncoder 학습 시, N-1년의 가중치를 초기 가중치로 적용하며, 클러스터링(K-Means) 과정에서도 N-1년의 Cluster Center를 초기값으로 지정하여 시계열적 안정성을 높혔다. 단, 초기 연도의 군집 결과는 참조할 N-1년의 데이터가 없으므로 초기 가중치를 별도로 설정하지 않았다.

[Figure 1] AutoEncoder Network Structure (Encoding/Decoding Network with Latent Space)



[Figure 2] Urban Engine FrameWork



DEC 모델의 적용은 크게 두가지 과정으로 구분할 수 있다. 우선 모델의 학습에 적용되는 Deep AutoEncoder 구조는 Encoder - LatentSpace - Decoder 구조로 이루어진 신경망 네트워크이다 (Figure 1 참조). Encoder에서 LatentSpace(잠재공간)을 통해 Decoder로 데이터가 다시 복원되는 과정을 거치며, 데이터가 복원되는 정도를 Loss로 하여 모델이 평가된다. 즉, 인허가 발생지역별로 구축된 데이터셋을 AutoEncoder를 통해 복원하는 과정에서 데이터의 복잡한 구조가 학습되게 되며, 복원되는 정도가 정확할 수록 신경망이 데이터의 특성을 잘 반영하게 된다. 따라서, AutoEncoder 자체가 복잡한 사회현상이 신경망에 적합되고 이를 기반으로 여러 데이터적 해석을 수행할 수 있는 것이다.

다음으로, AutoEncoder에서 도출되는 저차원의 Latent Space를 Input으로 하여, 실제 군집분석이 수행되는 과정이다. 일반적으로 군집분석에서 지나치게 높은 차원의 데이터를 사용하는 경우 “차원의 함정”에 빠지는 등 군집의 정확도가 떨어지고 사회 현상을 제대로 설명할 수 없는 문제가 종종 발생한다. 따라서, 차원축소 등의 방법을 적용하기도 하는데 데이터의 복잡한 구조가 손실될 수 있다는 위험성 또한 존재하는 방법론이다. 본 엔진에서는 이러한 문제점을 극복하기 위하여, AutoEncoder 학습 과정에서 도출되는 Latent Space의 Vector를 K-Means의 Input으로 군집분석을 수행하였다. AutoEncoder의 학습 과정에서 LatentSpace의 벡터 및 가중치들이 적절히 업데이트 되므로, LatentSpace 자체가 복잡한 네트워크를 간략히 설명해줄 수 있는 효과적인 도구로서 작용 할 수 있다. 이와 같은 개념을 활용하여, 본 엔진에서는 Deep Embedded K-Means를 적용하여 복잡한 데이터를 효과적으로 구분하는 방법론을 선택하였다.

특히, 시계열적 안정성을 확보하기 위하여 엔진의 최초 연도인 2018년을 제외하고, 2019년 모델부터 N-1년도의 AutoEncoder 모델의 가중치를 초기 가중치로 활용하게 된다. 이를 통하여 시계열적으로 안정적인 신경망 모델이 구축되었다. K-Means 과정에서 설정되는 Initial Clustering Center 또한 N-1년도의 Cluster Center Output을 활용하여, 기존의 군집 구조와 상이한 결과가 도출되는 것을 방지하였다. 특히, 이와 같은 방법을 통해서 군집의 갯수를 고정하고 초기에 설정된 군집의 특성이 시간이 지나서도 어느정도 유지되는 것을 기대할 수 있는 장점이 있다. 또한, 도시별 거시적 특징을 반영하기 위하여 시도별, 연도별 모델이 각각 구축되어, 군집내 동질성 및 군집 간 이질성이 최대한 확보될 수 있도록 하였다.

하이퍼 파라미터 튜닝 및 신경망 구조 설계, 활성화 함수 조

정 및 과적합 방지 등의 방법론은 모델의 학습 과정을 통해 도출되는 Loss 및 Validation Loss값을 바탕으로 경험적으로 모델 구축 및 최적화되었다.

4) Engine FrameWork

우선, Construction Action Detector에서 의미하는 “개별 건축 (개발) 행위”는 “특정 연도”에 인허가된 건축행위를 말한다. 인허가 연도를 기준 시점으로 하므로, 집계되는 데이터 또한 해당 연도에서의 속성정보를 기반으로 한다. 본 엔진에서의 구체적인 개발 프레임워크는 [Figure 2]와 같다. 해당 과정을 거쳐 최종적인 Engine이 개발되었다. 보다 구체적인 방법론에 대한 세부 설명은 Urban Engine 자산 관리 차원에서 공개하지 않으며, 다음 사항들에 대해서만 개략적으로 설명하였다.

[Table 1]을 기반으로 가공된 데이터들을 집계하는 기초 단위는 개별 건축 인허가 행위가 발생한 반경 250m로 설정되었으며, 50m Cell별 특성을 반경 250m 기반으로 집계하여, 인허가 발생 Cell별 데이터가 집계되었다. 즉, 건축 인허가가 이루어진 50m Cell별로 반경 250m 영향권의 다양한 속성정보들이 집계된 형태로 데이터를 관리하였다.

다음으로, Deep Clustering AI Engine의, Deep Embedded Clustering 모델을 활용하여, 개별 건축 행위를 유형화 하는 과정을 거쳤다. 연도별, 시도별 모델을 구축하고 건축 행위를 유형화하였으며, 이에 대한 세부적인 사항은 4-3)에 상세히 기술하였다. 추가적으로 최종 도출된 결과를 기반으로 유형화 결과에 대한 해석을 용이하게 하는 몇 가지 방법들이 적용되었다. ChatGPT v4o 모델을 활용하여, 시도별, 개별 건축행위 유형별 고유한 도시 특성들을 알기쉬운 스토리텔링 텍스트 형태로 기술하게 하였다. 또한, 군집을 설명 할 수 있는 핵심 변수들을 추출하고, 시각적으로 군집의 특성을 손쉽게 파악할 수 있는 시각화 결과물을 도출하였다. 이와 같은 요소들을 바탕으로 Urban Engine 이용자들이 유형화 결과에 대해 보다 쉽게 이해할 수 있는 장치를 마련하였다.

마지막으로, 엔진에서 도출된 결과물들은 Urban Engine 시각화 Tool을 기반으로 개발되었다. 해당 자료들은 DashBoard 형태로 시도 단위로 제공되며, 시군구별로 선택하여 다양한 정보들을 확인할 수 있는 구조를 가지고 있다. 개별 건축행위의 유형은 인구, 종사자 수, 건축행위 수 등과 같은 개발 환경 및 도시 환경을 반영하는 주요한 변수들과 군집의 특성을 통계적으로 구분할 수 있는 영향력이 높은 핵심 변수들을 바탕으로 시각화 결과물을 제공하였다.

5. Results

Construction Action Detector는 개별 건축 행위가 발생하는 지역 및 인근 지역의 특성을 기반으로 시군구의 건축 행위에 대해 체계적이고 지속적으로 식별하는 것을 목적으로 하고 있다. 또한, 추후에 개발될 예정인 도시 개발의 잠재력을 실제 측정하고, 도시개발 가능성성이 높은 지역을 식별하는 엔진에서 기초적으로 활용되는 것이 주 목적이다. 따라서, 제공되는 데이터 셋 자체가 기초적으로 활용될 수 있는 특성을 가지고 있다. 미시적인 결과는 일부 핵심 변수에 대해 세부적으로 시각화를 통해 제공하며, 대부분의 데이터는 시군구와 군집 유형 단위로 집계된 형태로 제공되어 시군구의 전체적인 특성을 파악하는데 보다 적합하다. 소지역 단위의 세부적인 개별 건축 개발 특성들은 추후 별도의 서비스 형태로 제공될 예정이다.

1) Data Descriptions

현재까지 Engine에 포함 및 개발된 데이터의 세부 정보는 <Table 2>와 같다. <Table 1>의 원천 데이터를 기반으로 파생된 다양한 데이터셋이 개발되어 Engine 내에 포함되어 있다. 데이터 칼럼에 대한 세부적인 정보는 필요에 따라, 가변적으로 업데이트 될 예정이다. 특히, 교통 및 기반시설 등과 관련된 데이터들은 Urban Transportaion Engine, Urban Infra Engine 등이 개발됨에 따라 기초 데이터 및 파생 데이터 등이 연계되어 추가될 예정에 있다. 따라서, 모델의 정확도와 설명력 또한 점차 높아질 예정이며, 일부 건축행위에 핵심적인 영향을 끼치는 데이터가 아직 미개발 되어 있는 점이 차차 보완될 것이다. 또한, 공개하는 데이터 수준으로는 시군구 단위를 최소로 <Table 2>의 변수들을 집계하여 제공한다.

[Table 2] Construction Action Detector Data Descriptions (계속)

Data Type	Data Name (Eng.)	Data Name (Kor.)	Descriptions
기초정보 (9)	base_year	데이터 기준연도	2018~2023년 생산
	const_id	개별 건축 행위 발생 지역 ID	지역 고유 ID, 시군구(5) + 고유번호(4)
	sido_cd	시도코드	공개용 데이터 수준
	cty_cd	시군구코드	
	grid_id	국가표준격자(50m) ID	내부용(미공개) 데이터 수준
	x_axis	건축 행위 발생지역 X좌표	
	y_axis	건축 행위 발생지역 Y좌표	
	cluster_id	건축 행위 유형 구분 코드	시도별 상이한 Cluster 정보
	bdtype_nm	건축물 유형 정보	인허가 대장 상 건축물 유형 (단독주택, 공동주택, 제1, 2종근린생활시설 등)
건축 행위 정보 (2)	const_cnt	건축 행위 발생 횟수	
	effect_num	건축 행위 영향권 누적도	250m 영향권 영역 중첩 정도
인구정보 (3)	pop_female	여성인구 수	격자 단위 인구 데이터
	pop_male	남성인구 수	
	pop_tot	총 인구 수	
사업체 정보(2)	tot_job_buz	총 사업체 수	
	tot_job_pop	총 종사자 수	
토지 지목 정보 (22)	lot_1_area	지목(전) 면적	과수원, 목장용지가 포함된 면적
	lot_2_area	지목(답) 면적	
	lot_5_area	지목(임야) 면적	
	lot_6_area	지목(광천지) 면적	염전 용지가 포함된 면적
	lot_8_area	지목(대지) 면적	
	lot_9_area	지목(공장용지) 면적	
	lot_10_area	지목(학교용지) 면적	
	lot_11_area	지목(주차장) 면적	
	lot_12_area	지목(주유소용지) 면적	
	lot_13_area	지목(창고용지) 면적	
	lot_14_area	지목(도로) 면적	
	lot_15_area	지목(철도용지) 면적	
	lot_17_area	지목(하천) 면적	유지 용지가 포함된 면적
	lot_18_area	지목(구거) 면적	제방, 수도용지가 포함된 면적
	lot_20_area	지목(양어장) 면적	

Data Type	Data Name (Eng.)	Data Name (Kor.)	Descriptions
토지 지목 정보 (22)	lot_22_area	지목(공원) 면적	
	lot_23_area	지목(체육용지) 면적	
	lot_24_area	지목(유원지) 면적	
	lot_25_area	지목(종교용지) 면적	
	lot_26_area	지목(사적지) 면적	
	lot_27_area	지목(묘지) 면적	
	lot_28_area	지목(잡종지) 면적	
지가정보 (1)	land_price	평균 공시지가	250m 영향권 내 필지의 평균 공시지가
건축물 특성정보 (22)	tot_build_cnt	총 건물 개수	
	tot_avg_flr	건물 평균 층수	250m 영향권 내 건축물 평균 층수
	tot_max_flr	건물 최고 층수	250m 영향권 내 건축물 최고 층수
	tot_grid_bcr	격자 건폐율	영향 권역 내 포함된 국가격자의 평균 건폐율
	tot_grid_far	격자 용적률	영향 권역 내 포함된 국가격자의 평균 용적률
	tot_grs_flr_area	총 연면적	반경 250m 내 건축물의 총 연면적
	dtd_house_area	단독주택 연면적	
	mult_house_area	공동주택 연면적	
	nlf1_facil_area	제1종 균린생활시설 연면적	
	nlf2_facil_area	제2동 균린생활시설 연면적	
	cult_facil_area	문화 및 집회시설 연면적	
	sale_facil_area	판매 및 영업시설 연면적	
	medi_facil_area	의료시설 연면적	
	edwf_facil_area	교육연구 및 복지시설 연면적	
	sppt_facil_area	운동시설 연면적	
	offc_facil_area	업무시설 연면적	
	lodg_facil_area	숙박시설 연면적	
	amus_facil_area	위락시설 연면적	
	fact_facil_area	공장시설 연면적	
	stor_facil_area	창고시설 연면적	
	pubc_facil_area	공공용시설 연면적	
	tour_facil_area	관광휴게시설 연면적	
기반시설 특성정보 (12)	road_rank_0	총 도로 면적	
	road_rank_1	도로면적 (소로)	도로폭 12m 미만의 도로면적 총합
	road_rank_2	도로면적 (중로)	도로폭 12m 이상 20m 미만의 도로면적 총합
	road_rank_3	도로면적 (대로)	도로폭 25m 이상 35m 미만의 도로면적 총합
	road_rank_4	도로면적 (광로)	도로폭 40m 이상의 도로면적 총합
	park_all	총 공원 면적	
	park_2.5kun	공원 면적 (2,500㎡ 이하)	
	park_2.5k_10k	공원 면적 (2,500㎡ 이상, 1만㎡ 이하)	
	park_10k_30k	공원 면적 (1만㎡ 이상, 3만㎡ 이하)	
	park_30k_100k	공원 면적 (3만㎡ 이상, 10만㎡ 이하)	
	park_100kup	공원 면적 (10만㎡ 이상)	
	builtup_area	시가화 면적	지목과 건축물 데이터를 결합하여 계산된 시가화 면적
등록사업체 운영정보 대분류 (32)	animal_oper_cnt	동물 업종 운영업체 수	
	animal_oper_intv	동물 업종 평균 영업 기간	
	animal_regi_cnt	동물 업종 신규 등록 업체 수	
	animal_close_cnt	동물 업종 폐업 업체 수	
	culture_oper_cnt	문화 업종 운영업체 수	
	culture_oper_intv	문화 업종 평균 영업 기간	
	culture_regi_cnt	문화 업종 신규 등록 업체 수	
	culture_close_cnt	문화 업종 폐업 업체 수	
	food_oper_cnt	식품 업종 운영업체 수	
	food_oper_intv	식품 업종 평균 영업 기간	

인허가 데이터를 기반으로 재분류한 대분류 업종

Data Type	Data Name (Eng.)	Data Name (Kor.)	Descriptions
등록사업체 운영정보 대분류 (32)	food_regi_cnt	식품 업종 신규 등록업체 수	인허가 데이터를 기반으로 재분류한 대분류 업종
	food_close_cnt	식품 업종 폐업업체 수	
	health_oper_cnt	건강 업종 운영업체 수	
	health_oper_intv	건강 업종 평균 영업 기간	
	health_regi_cnt	건강 업종 신규 등록업체 수	
	health_close_cnt	건강 업종 폐업업체 수	
	life_oper_cnt	생활 업종 운영업체 수	
	life_oper_intv	생활 업종 평균 영업 기간	
	life_regi_cnt	생활 업종 신규 등록업체 수	
	life_close_cnt	생활 업종 폐업업체 수	
	natural_oper_cnt	자원환경 업종 운영업체 수	
	natural_oper_intv	자원환경 업종 평균 영업 기간	
	natural_regi_cnt	자원환경 업종 신규 등록업체 수	
	natural_close_cnt	자원환경 업종 폐업업체 수	
	restaurant_oper_cnt	음식점 업종 운영업체 수	
	restaurant_oper_intv	음식점 업종 평균 영업 기간	
	restaurant_regi_cnt	음식점 업종 신규 등록업체 수	
	restaurant_close_cnt	음식점 업종 폐업업체 수	
	etc_oper_cnt	기타 업종 운영업체 수	
	etc_oper_intv	기타 업종 평균 영업 기간	
	etc_regi_cnt	기타 업종 신규 등록업체 수	
	etc_close_cnt	기타 업종 폐업업체 수	
등록사업체 운영정보 중분류 (44)	anim_groo	동물미용업 운영업체 수	인허가 데이터 상 업종 분류 레벨 2단계에 해당하는 업종
	anim_hosp	동물병원 운영업체 수	
	anim_live	가축사육업 운영업체 수	
	anim_phar	동물약국 운영업체 수	
	anim_prod	동물생산업 운영업체 수	
	cult_inet	인터넷컴퓨터게임시설제공업 운영업체 수	
	cult_kara	노래연습장업 운영업체 수	
	cult_lodg	숙박업 운영업체 수	
	cult_farm	농어촌민박업 운영업체 수	
	etc_bath	목욕장업 운영업체 수	
	etc_door	방문판매업 운영업체 수	
	etc_sale	통신판매업 운영업체 수	
	etc_store	대규모점포 운영업체 수	
	etc_toba	담배소매업 운영업체 수	
	food_bake	제과점영업 운영업체 수	
	food_dist	유통전문판매업 운영업체 수	
	food_hlth	건강기능식품일반판매업 운영업체 수	
	food_inst	즉석판매제조가공업 운영업체 수	
	food_mass	집단급식소 운영업체 수	
	food_meat	축산판매업 운영업체 수	
	food_vend	식품자동판매기업 운영업체 수	
	heal_clin	의원 운영업체 수	
	heal_dent	치과기공소 운영업체 수	
	heal_hosp	병원 운영업체 수	
	heal_medi	의료기기판매(임대)업 운영업체 수	
	heal_opti	안경업 운영업체 수	
	heal_otc	안전상비의약품 판매업소 운영업체 수	
	heal_phar	약국 운영업체 수	
	life_ad	옥외광고업 운영업체 수	
	life_beau	미용업 운영업체 수	

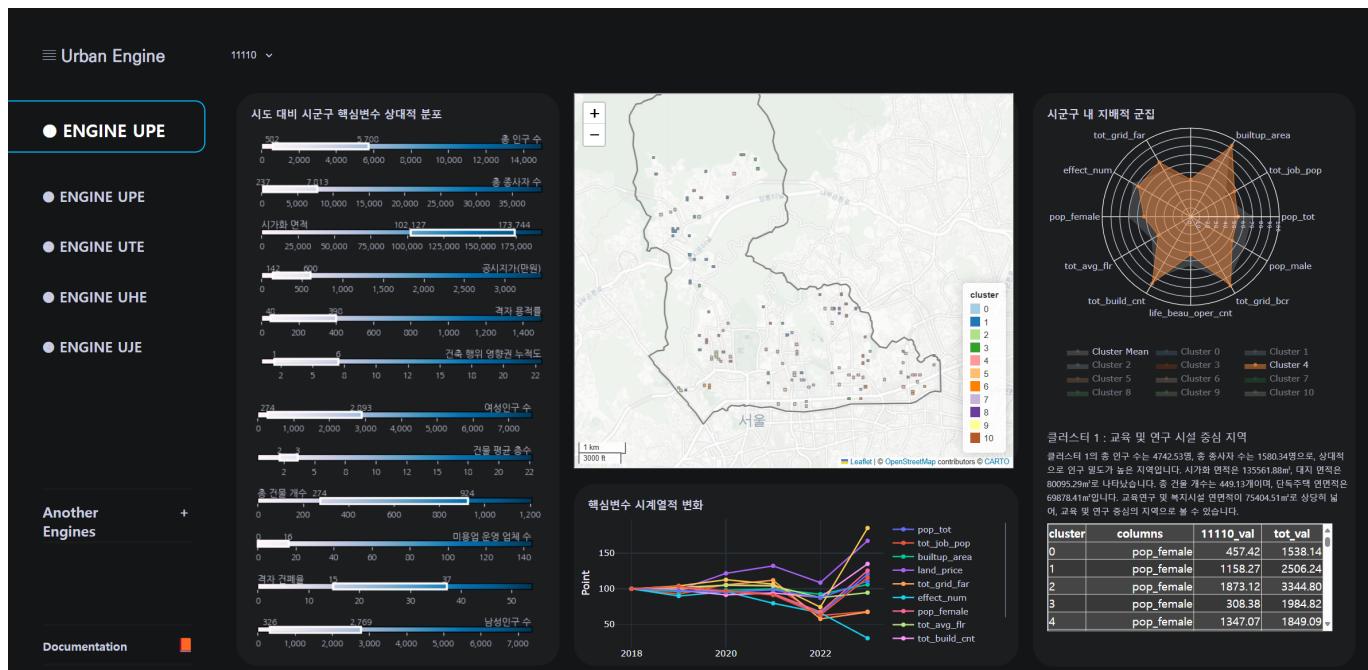
Data Type	Data Name (Eng.)	Data Name (Kor.)	Descriptions
등록사업체 운영정보 중분류 (44)	life_bill	당구장업 운영 업체 수	인허가 데이터 상 업종 분류 레벨 2단계에 해당하는 업종
	life_golf	골프연습장업 운영 업체 수	
	life_gym	체력단련장업 운영 업체 수	
	life_laun	세탁업 운영 업체 수	
	life_mart	대형마트 운영 업체 수	
	life_publ	출판사 운영 업체 수	
	natu_air	대기오염물질배출시설설치사업장 운영 업체 수	
	natu_bldg	건물위생관리업 운영 업체 수	
	natu_oil	석유판매업 운영 업체 수	
	natu_trash	쓰레기종량제봉투판매업 운영 업체 수	
	rest_club	유흥주점영업 운영 업체 수	
	rest_fast	휴게음식점 운영 업체 수	
	rest_gen	일반음식점 운영 업체 수	
	rest_loung	단란주점영업 운영 업체 수	

2) 시각화 결과

구축된 엔진을 바탕으로 Urban Engine 시각화 Tool을 이용하여 실제 시각화 한 결과는 <Figure 3>와 같다. 효율적인 시각화를 위해 개별 건축 개발 행위와 연관성이 높은 기초적 도시특성에 해당하는 변수들과, 개별 건축 행위 유형을 잘 설명하는 핵심 변수들을 대상으로 시각화를 수행하였다. 기초적 도시특성으로는 총 인구 수, 총 종사자 수, 시가화 면적, 공시지가, 격자 용적률, 건축 행위 영향권 누적도로 설정하다. 건축 행위 유형을 설명하는 변수로는 시도별 군집을 구분하는데 큰 영향을 미치는 변수들을 자동으로 설정하였으며, 시도별로 서로 상이한 변수가 추출되었다.

시각화 대시보드는 기본적으로 시도별로 제공되며, 시도 내 특정 시군구를 선택할 수 있도록 설계되어 있다. 특정 시군구 가의 개별 건축 행위 발생 지역들이 시도 내에서 어느 정도의 수준을 갖는지를 Box-Range Plot Chart를 이용하여 확인할 수 있다. 연도별 주요 변수들의 변화 양상 또한, Percent Ratio를 사용하여 2018년을 100으로 지수화하여 제공하고 있다. 개별 건축 행위 발생 지역들을 시도별로 군집화하고, 군집 내 세부적인 설명과 데이터 또한 제공하여 특정 지역에서 발생한 건축 행위가 어떠한 유형에 속해있고, 어떤 특징을 가지고 있는지 쉽게 확인할 수 있는 장점이 있다. 이를 통해, 시군구 내 어떤 유형에서 건축 행위가 다수 발생하고 있는지를 일목요연하게 파악 할 수 있다.

[Figure 3] Construction Action Detector 시각화 Dashboard



6. Conclusions

본 연구는 Urban Planning Engine 프로젝트의 일환으로 개발된 Construction Action Detector의 설계 및 구현에 대해 다루었다. 본 엔진의 개발 목적은 개별 건축 인허가 행위의 추세와 이를 둘러싼 지역 특성을 체계적으로 탐색하고 분석함으로써, 도시 개발의 미시적 단위에서 발생하는 건축 행위를 보다 효과적으로 이해하고자 하는 것이다. 본 엔진의 성과와 의의는 다음과 같다.

첫째, 본 엔진은 다양한 공공 데이터를 통합한 엔진 개발을 통해, 개별 건축 개발 행위와 관련된 다양한 측면을 다루고 복잡한 양상을 추적하고 있다. 건축물 대장, 인구, 종사자, 전자지도, 인허가 데이터 등 다양한 소스의 공공 데이터를 활용하여, 건축 행위가 발생한 지역의 특성을 정밀하게 파악하였다. 특히, 건축 인허가 데이터를 기반으로 시군 구별 건축 행위의 시계열적 탐색 및 공간 분포를 식별함으로써, 각 지역의 개발 양상과 패턴을 명확히 이해할 수 있었다.

둘째, 본 엔진은 Urban Engine의 내부 기술 Core 엔진인 Deep Clustering AI Engine을 적용하여, 개별 건축 행위가 발생한 지역에 대한 유형화를 수행하였다. 비지도 학습 기법인 Deep Embedded Clustering (DEC) 모델을 활용하여, 데이터의 군집화를 통해 지역별 건축 행위의 특징을 도출하였다. DEC 모델의 활용을 통해 시계열적 안정성을 확보 할 수 있었으며, 이는 연도별 군집의 일관성을 유지하는데 기여하였다. 이러한 군집화 결과는 향후 도시 개발 잠재력을 평가하는 데 중요한 기초 자료로 활용될 수 있다. 또한, Deep Clustering AI Engine 초기 모델의 실제 활용 버전으로서, 향후 AI 엔진에 대한 활용성을 기대할 수 있다.

셋째, 본 연구는 건축 행위 탐지기의 시각화 도구를 통해 분석 결과를 보다 쉽게 이해할 수 있도록 하였다. 시각화 대시보드는 시도별, 시군구별로 제공되며, 특정 지역에서 발생한 건축 행위의 유형과 특성을 직관적으로 파악할 수 있도록 설계되었다. 이를 통해 이용자들은 도시 개발의 미시적 단위에서 발생하는 다양한 건축 행위의 특징을 보다 명확하게 이해하고, 이를 바탕으로 효과적인 도시 개발 전략을 수립할 수 있을 것이다.

Construction Action Detector는 여러 장점을 가지고 있으나, 몇 가지 한계점도 존재한다. 첫째, 현재 엔진에 포함된 데이터는 주로 건축 인허가 데이터와 인구, 종사자 데이터에 한정되어 있어, 교통 및 기반 시설 등 다른 중요한 변수들은 포함되지 않았다. 향후 Urban Planning Engine 이외의 다양한 도메인에서의 Engine이 개발될 때, 양질의 변

수들이 추가되어 개선된 결과물들이 업데이트 되어야 한다. 둘째, 본 엔진의 Deep Clustering AI 모델은 초기 버전으로 추가적인 활용을 위해서는 모델 개선 및 검증과정이 지속적으로 필요하므로, 일부 결과들이 현상에 맞지 않을 가능성이 일부 존재한다. 이와 같은 일부 오류들은 엔진의 개선에 따라 지속적으로 업데이트 되어 고도화될 예정에 있다.

결론적으로, 본 연구에서 개발된 건축 행위 탐지기는 도시 개발의 미시적 단위에서 발생하는 건축 행위를 체계적으로 이해하고 분석할 수 있는 유용한 도구로서, 향후 도시 개발 전략 수립에 중요한 기초 자료로 활용될 수 있을 것이다. 지속적인 데이터 업데이트와 모델 고도화를 통해, 더욱 정교하고 효과적인 도시 계획 지원 도구로 발전할 것으로 기대된다.