

동시출현단어 분석을 활용한 건설 분야 인공지능 적용에 관한 지적구조 분석

김선겸*¹, 원지선*², 신재영**

Domain Analysis on the Field of Artificial Intelligence of Construction by Co-Word Analysis

Sun-Kyum Kim*¹, Jisun Won*², and Jaeyoung Shin**

본 연구는 본 연구는 과학기술정보통신부 한국건설기술연구원 연구운영비지원(주요사업)사업으로 수행되었습니다
(20230117-001, 미래 건설산업 견인 및 신시장 창출을 위한 스마트 건설기술 연구)

요 약

최근 인공지능에 대한 관심이 증가함에 따라, 건설 분야에서 인공지능을 적용한 최신 연구들을 체계적으로 검토하고 요약하여 과거의 연구들에 대한 분석 및 향후 적용 가능한 제언 등을 제시할 필요가 있다. 본 연구는 동시출현단어 분석을 통해 건설 분야의 인공지능 적용에 대한 최신 연구 경향을 반영하는 지적구조를 제시하여 주제영역 구성을 나타내는 것이며, 2018년 1월 1일부터 2022년 12월 31일까지 최근 5년간의 Web of Science 데이터베이스로부터 문헌 데이터를 수집하고 동시출현단어를 활용하여 분석한다. 이를 위하여 건설 분야의 인공지능 적용에 관한 지적구조 규명을 위해 네트워크 분석, 중심성 분석, 클러스터 분석을 통해 각 키워드들 간의 상관관계에 따른 지적구조를 제시하였다.

Abstract

As interest in artificial intelligence has recently increased, it is necessary to systematically review and summarize the latest studies applying artificial intelligence in the field of construction, analyze past studies, and present applicable suggestions for the future. This study uses co-occurring word analysis to present a domain structure that reflects the latest research trends on the application of artificial intelligence to the construction field to reveal the composition of the subject area, from January 1, 2018 to December 31, 2022. Literature data collected is from Web of Science database and analyzed using co-occurring words. To this end, in order to identify the domain structure related to the application of artificial intelligence in the construction field, the domain structure according to the correlation between each keyword was indicated through network analysis, centrality analysis, and cluster analysis.

Keywords

artificial intelligence, deep learning, domain analysis, co-word analysis, network analysis

* 한국건설기술연구원 수석연구원(*¹ 교신저자)
- ORCID¹: <https://orcid.org/0000-0002-4610-1961>
- ORCID²: <https://orcid.org/0000-0002-3690-8470>
** 한국건설기술연구원 전임연구원
- ORCID: <https://orcid.org/0000-0002-5917-0472>

· Received: Jun. 07, 2023, Revised: Jul. 05, 2023, Accepted: Jul. 08, 2023
· Corresponding Author: Sun-Kyum Kim
Dept. of Future & Smart Construction Research
Korea Institute of Civil Engineering and Building Technology
Tel.: +82-31-995-0962, Email: sunkyumkim@kict.re.kr

1. 서 론

인공지능은 인간의 지능과 유사한 방식으로 복잡한 데이터를 대량으로 처리할 수 있고, 이를 통해 다양한 문제 해결 및 실시간 의사결정을 가능케 하는 기술이다[1]. 머신러닝과 딥러닝은 인공지능의 포함되는 개념으로 머신러닝은 컴퓨터가 프로그래밍 되지 않고 학습하는 방법을 연구하는 분야이며, 딥러닝은 컴퓨터가 과거 경험을 통해 학습할 수 있도록 하는 분야이다[2].

최근 건설 산업은 스마트건설 기조에 발맞추어 인공지능의 적용을 시도해오고 있으며, 이의 긍정적 결과로 건설 분야의 인공지능 연구는 더욱 탄력을 받고 있다. 예컨대, 이상 탐지, 손상 감지, 모니터링 등을 비롯한 데이터 인식, 처리 및 의사결정을 위해 머신러닝과 딥러닝을 활용하였다[3][4].

인공지능에 대한 관심이 증가함에 따라, 건설 분야에서 인공지능을 적용한 최신 기술들을 체계적으로 검토하고 요약하여 과거의 연구들에 대한 분석 및 향후 적용 가능한 제언 등을 제시할 필요가 있다.

관련 문헌에 대한 지적구조 분석을 통해 건설 분야 인공지능 적용에 대한 이해가 가능하다. 지적구조 분석의 대표적인 방법인 동시출현단어 분석은 문헌을 구성하는 제목, 초록, 키워드, 텍스트 전문을 이용하여 주제영역으로부터 하위분야 및 분야의 패턴의 분석을 통해 해당 분야의 발전 방향을 예측할 수 있는 근거가 될 수 있다[5].

본 연구의 목적은 동시출현단어 분석을 통해 건설 분야의 인공지능 적용에 대한 최신 연구 경향을 반영하는 지적구조를 제시하고 주제영역 구성을 나타내는 것이다. 이를 위하여 건설 분야의 인공지능 적용에 관한 논문 데이터를 수집하고 분석하였다. 2018년부터 2022년까지 최근 5년간 발간된 건설 분야의 인공지능 적용에 관한 논문을 Web of Science 데이터베이스로부터 수집하였다. 수집된 총 810건의 논문 중 제목과 초록에서 추출된 키워드는 총 4,197개였다. 키워드 관계 네트워크의 세부 주제 영역을 파악하기 위해 추출된 키워드를 통해 네트워크 분석을 수행을, 관련 키워드들이 얼마만큼 서로간에 영향을 주는지 중심성 분석을 수행하였다. 마지막으

로 클러스터 분석을 수행하고, 이를 다차원축적지도를 통해 건설 분야의 인공지능 적용에 관한 전체적인 주제영역의 흐름 및 구성을 제시하였다. 본 연구의 결과는 최근 건설 분야의 인공지능 적용에 대한 최신 연구 경향 및 주제영역에 대한 정보 제공이며, 이를 통한 연구의 발전방향에 대하여 도움이 될 수 있을 것으로 기대한다.

본 논문은 2장에서 관련된 지적구조를 위한 데이터 추출 및 행렬 작성하며, 건설 분야의 인공지능 적용 관련 문헌에 관한 지적구조를 위한 데이터 추출 및 행렬 작성을 설명한다. 3, 4, 5장은 이를 토대로 한 네트워크 및 클러스터, 다차원축적지도를 통한 지적구조 분석결과를 보여주며, 마지막 6장으로 결론과 의의를 논하는 것으로 마무리한다.

II. 데이터 추출 및 행렬 작성

2.1 지적구조 분석

건설 분야의 인공지능 적용에 대한 지적구조 분석을 위해서는 그림 1과 같이 데이터의 추출과 가공 및 분석이 필요하다. 데이터 수집과 키워드 추출을 위하여 Web of Science 소속 저널에 게재된 문헌을 수집하였다. Web of Science에 소속된 저널 범주를 공학 토목(Engineering civil)과 건설 건축 기술(Construction building technology)로 제한하였고, 기간은 2018년 1월 1일부터 2022년 12월 31일까지의 데이터를 수집하였다. 분석 대상이 되는 건설 분야의 인공지능 적용 관련 문헌들을 추출하기 위해 첫 번째 키워드를 ‘construction’로, ‘artificial intelligence’와 ‘deep learning’ 키워드를 ‘or’로 활용한 두 번째 키워드로 하여 주제(Topic) 검색을 실시하였다. 검색 결과, 건설 분야 인공지능 관련 연구 논문 729건과 Review article 논문 76건, 편집자료 4건, 프로시딩 논문 1건, 총 810건이 수집되었다. 이를 네트워크 시각화 소프트웨어인 CiteSpace 6.2를 통해 수집된 데이터로부터 의미 있는 단어들을 추출하기 위해 전처리(Preprocess) 분석을 통해 최종적으로 분석에 필요한 데이터 셋을 제작하였다.

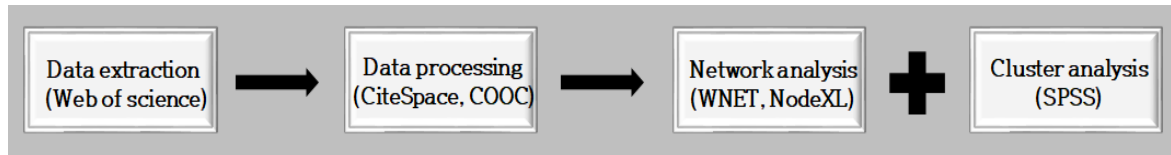


그림 1. 지적구조 분석 과정
Fig. 1. Domain analysis process

제작된 데이터셋을 활용하여 COOC ver 0.4 [6]을 통해 동시출현단어를 추출하고, 이를 코사인 유사도 및 피어슨 상관계수를 활용하여 단어별 유사도를 구하고, 단어 간 어떠한 관계를 가지고 있는지 파악하였다. 이후, 병렬 최근접 이웃 클러스터링 알고리즘(PNNC)를 활용하여 WNET ver 0.4[7]를 통해 중심성 분석 통해 네트워크 분석을, Nodexl[8]를 통해 시각화를 수행하였다. 또한 SPSS ver. 29.0을 통해 클러스터 분석을 하였고, 건설분야의 인공지능 적용 영역에서의 흐름을 확인하였다.

2.2 키워드 선정

본 연구에서는 수집된 810건의 문헌들을 CiteSpace 6.2을 활용하여 제목과 초록에서 키워드 명사구와 빈도수를 추출하였다. 상세한 주제 영역 구성을 위하여 명사구당 2-4개 사이를 범위 개수로 지정하였으며, 초록이 없는 경우, 제목에서만 명사구를 추출하였다. 이 810건의 문헌들에서 추출된 명사구는 총 4,197개(문헌 당 5.18개)이었다. 이 중 분석을 용이하게 하기 위하여 불용어를 제거하고, 규모를 축소하여 자주 출현한 명사구 363개의 리스트를 추출하였다. 나아가 복잡성을 고려하여 너무 많은 키워드를 포함하지 않도록, 빈도수 전체의 70%에 해당하는 8회 이상 등장한 키워드 83개로 구성된 리스트를 완성하였다.

키워드의 정규화를 위하여 단수와 복수의 명사구 중 ‘point clouds’ 및 ‘point cloud’ 또는 ‘construction sites’ 및 ‘construction site’와 같은 경우 같이 복수형태가 주로 사용되는 명사구인 ‘point clouds’와 ‘construction sites’로 모두 변경하였으며, ‘neural network’ 및 ‘neural networks’와 같은 경우 단수 형태를 주로 쓰는 ‘neural network’로 통일하였다. 또한 ‘learning algorithms’ 또는 ‘learning methods’와 같이 비슷한 단어이고 뜻도 동일한 단어들도 빈도수가

보다 높은 ‘learning algorithms’로 치환하였으며, ‘construction safety’ 및 ‘safety management’와 같이 단어는 다르나 뜻이 동일한 경우에도 ‘construction safety’와 같이 빈도수가 높은 키워드로 치환하였다. 그 외에 ‘american society’와 같은 관련이 없는 키워드는 삭제하였다.

앞서 추출된 키워드는 하나의 논문에서 중복 출현된 단어의 빈도수가 모두 포함하여 합산되어 있지만, 연구의 분석에서 필요한 키워드는 각 논문에서 한번씩만 출현한 횟수이기 때문에 단어 빈도와 문헌 빈도의 값의 치우침에 따른 복잡성을 방지하기 위하여 단어 빈도에 대한 문헌 빈도 값인 TF/IDF 값이 백분위 90%인 47.04 이상인 최종 키워드 리스트를 선정하였다. 이에 대한 최종 54개의 리스트는 표 1과 같다.

건설 분야 인공지능 적용을 대표로 하는 키워드는 빈도수에 따라 deep learning(1), artificial intelligence(2), neural network(3), convolution neural network(4), machine learning(5), computer vision(6), deep learning algorithm(7)로 인공지능과 딥러닝의 대표 키워드이지만 실질적으로 construction industry(8), construction site(9), building information modeling(11), concrete structures(17), compressive strength(19), crack detection(24), construction management(40), construction equipment(43) 등 키워드를 통해 건설분야의 다양한 측면에서 인공지능을 활용하고 있다는 것을 알 수 있다.

2.3 동시출현단어 행렬 작성

동시출현단어 행렬은 가공 방법에 따라 네트워크의 형태의 달라짐으로 인하여 키워드 간 연관성 측정을 위해 유사도를 측정하는 방법인 코사인 유사도와 피어슨 상관계수를 활용하였다.

표 1과 같이 54개의 키워드가 선정된 문헌 810건의 정보를 엑셀에 모두 입력하고 COOC ver 0.4를 이용하여 동시출현단어 행렬을 작성하였다.

엑셀시트 첫 번째 열에 각 문헌 번호, 두 번째 열에 제목과 초록에서 추출된 키워드를 입력하여 추출 작업을 끝낸 후, 필요한 최종 키워드를 포함한 행을 남겨서 최종 분석 대상 출현 단어 리스트를 완성하였다. COOC ver 0.4를 통해 최종 키워드 출현빈도 정보 리스트, 최종 키워드 리스트, 동시출현단어 행렬을 파일로 만들었다. 이 동시출현단어 행렬은 정방대칭행렬이며, 54×54의 동시출현빈도와 대각선 간의 출현빈도로 구성된다. 이 후에 이 행렬 파일을 다시 입력하면 코사인 유사도계수와 피어슨 상관계수로 정규화한 행렬이 출력 된다[9].

코사인 유사도는 두 단어가 1에 가까울수록 유사도가 높고 0에 가까울수록 낮다. 코사인 유사도 분석결과, 1차 연관성 행렬인 코사인 유사도 행렬에서 유사도가 가장 높은 키워드 쌍은 ‘learning algorithm’-‘support vector machine’(0.32962)으로 나타났다. 그다음에는 ‘semantic segmentation’-‘point clouds’(0.30906), ‘learning algorithm’-‘random forest’(0.27937) 순으로 나타났다. 유사도가 0인 키워드 쌍을 제외하고, 유사도가 가장 낮은 것으로 나타난 키워드 쌍은 ‘artificial intelligence’-‘novel methods’(0.01292)로 나타났으며, 그다음으로 ‘artificial intelligence’-‘convolution neural network’(0.01355), ‘deep learning’-‘compressive strength’(0.01629) 순으로 매우 낮은 유사도를 보여주었다.

표 1. 최종 선정된 키워드 리스트

Table 1. Final selected keyword list

No.	Keyword	Frequency	No.	Keyword	Frequency
1	deep learning	348	28	model performance	23
2	artificial intelligence	172	29	feature extraction	22
3	neural network	130	30	random forest	21
4	convolution neural network	118	31	construction workers	21
5	machine learning	87	32	long short-term memory	18
6	computer vision	68	33	big data	18
7	deep learning algorithm	65	34	mean average precision	17
8	construction industry	64	35	f1 score	17
9	construction site	59	36	accurate prediction	16
10	deep learning model	52	37	prediction accuracy	16
11	building information modeling	44	38	transfer learning	16
12	previous studies	42	39	training data	16
13	point clouds	37	40	construction management	15
14	learning algorithm	34	41	large number	14
15	construction project	33	42	input data	14
16	semantic segmentation	31	43	construction equipment	13
17	concrete structure	31	44	hybrid model	13
18	future research	31	45	genetic algorithm	13
19	compressive strength	30	46	extreme gradient	11
20	novel method	29	47	systematic review	11
21	high accuracy	27	48	great significance	11
22	deep neural network	26	49	important role	10
23	civil engineers	26	50	automatic detection	10
24	crack detection	24	51	superior performance	10
25	recent years	24	52	great potential	9
26	natural language processing	23	53	sensitivity analysis	9
27	different types	23	54	support vector machine	8

피어슨 상관계수에 의해 산출된 값은 관계의 크기와 방향을 동시에 가지며, 범위는 -1에서 +1사이를 나타낸다. +1에 가까울수록 강한 긍정적 관계를 -1에 가까울수록 강한 부정적 관계를 뜻하며, 0은 두 키워드 간에 선형적인 관련성이 없음을 나타낸다. 피어슨 상관계수를 통해 분석한 결과, 가장 상관관계가 높은 값을 갖는 키워드 쌍은 ‘civil engineers’-‘construction worker’(0.63707)이며, 그다음으로는 ‘learning algorithm’-‘support vector machine’(0.61948), ‘semantic segmentation’-‘point clouds’(0.59877) 순으로 나타났으며, 정적인 높은 상관관계를 갖는다. 음의 상관관계가 가장 큰 키워드 쌍은 ‘construction project’-‘feature extraction’(-0.15721), ‘construction project’-‘crack detection’(-0.16903), ‘construction’-‘transfer learning’(-0.13789) 순으로 나타났으며, 경미한 음의 상관관계를 갖는다. 이밖의 음의 상관관계를 지닌 키워드 쌍은 총 634개였다.

III. 지적구조 분석결과

3.1 네트워크 생성

키워드간 관계를 조사하기 위하여 동시출현단어 분석을 통한 네트워크 시각화를 수행하며, 이 네트워크 관계에 따라 나뉘어진 클러스터로 인해 주제영역을 확인할 수 있다. 본 연구에서는 건설 분야의 인공지능 적용에 대한 지적구조를 분석하기 위해 피어슨 상관계수를 활용한 패스파인더 네트워크 알고리즘[10]을 적용하여 네트워크를 생성하였다. 그리고 패스파인더 네트워크상의 세부주제들을 명확하게 하기 위해 병렬 최근접 이웃 클러스터링 알고리즘[11]을 이용하였다. 네트워크 및 클러스터링 생성을 위해 WNET을 활용하였고, 시각화를 위해 NodeXL을 사용하였다. 키워드 54개에 관한 네트워크 분석 결과, 그림 2와 같이 PNNC 알고리즘으로 11개의 최적의 클러스터가 생성되었다. 그림의 명확성을 위해 클러스터의 각 세부 영역을 모두 병합하였다. 또한 각 클러스터에서 빈도수가 가장 높은 키

워드를 해당 클러스터의 주제명으로 부여하였으며, 이를 표 2에 표시하였다.

표 2. 클러스터별 대표 키워드

Table 2. Representative keywords by each cluster

No.	Representative keywords by each cluster
1	deep learning
2	artificial intelligence
3	neural network
4	deep learning algorithm
5	construction industry
6	deep learning model
7	civil engineers
8	compressive strength
9	computer vision
10	model performance
11	genetic algorithm

그림 2의 병합을 해제하여, 그림 3, 그림 4와 같이 각 11개의 클러스터에 속한 하위 주제영역을 나타내었고, 중심성 분석을 통해 전역중심성, 지역중심성, 매개중심성이 높은 키워드를 확인하였다. 키워드의 전역중심성이 높을수록 전반적으로 다수의 주변 키워드들과 연결되어 있는 키워드를 나타내며, 지역중심성이 높을수록 자기가 속한 클러스터 내에 하위 주제영역에서 영향력이 있는 키워드이며, 매개중심성이 높을수록 각 클러스터를 연결해주는 핵심중개자 역할을 한다.

3.2 네트워크 분석을 통한 지적구조 분석

3.1에서 생성된 네트워크는 빈도수를 통해 생성된 가중 네트워크이므로 중심성 분석을 하기 위해 가중 네트워크에서 활용하는 중심성 분석 척도[12]를 활용하였다. 삼각매개중심성과 평균연관성 측정을 통해 전역중심성이 높은 키워드를 확인하였으며, 이를 반영하여 그림 3과 같이 네트워크를 작성하였다.

16 동시출현단어 분석을 활용한 건설 분야 인공지능 적용에 관한 지적구조 분석

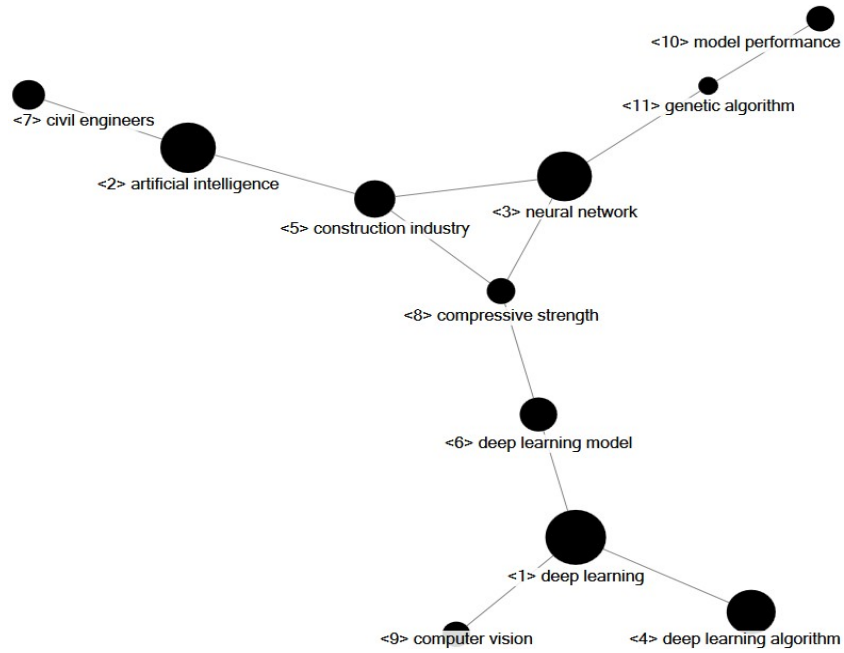


그림 2. 11개의 클러스터
Fig. 2. 11 Clusters

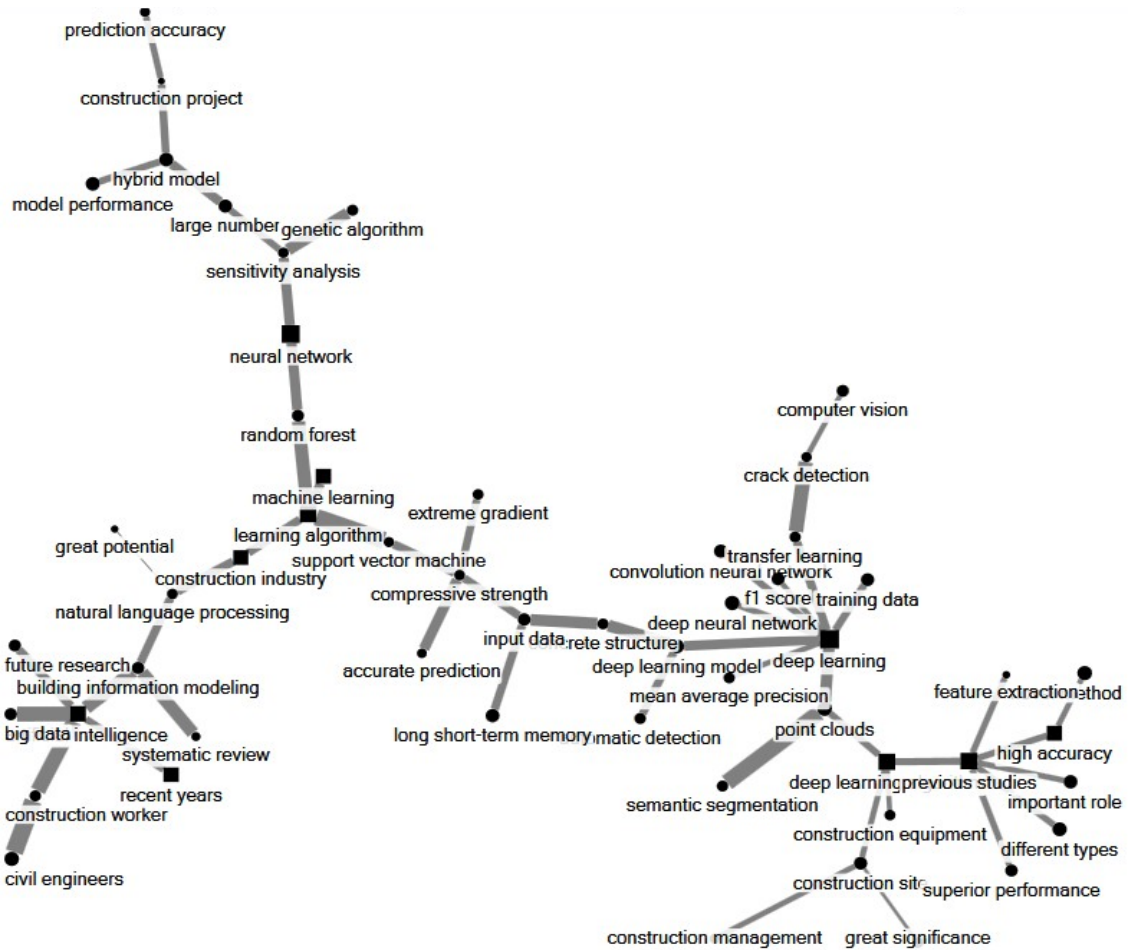


그림 3. 상대적 전역중심성에 의한 키워드 관계 네트워크
Fig. 3. Keyword relations network by global centrality

중심성은 네트워크를 구성하는 동시출현단어의 강도, 영향력 범위, 구조적인 입지를 반영한 값이다. 중심성은 네트워크를 구성하는 동시출현단어의 네트워크 구조의 핵심을 차지하는 위치, 영향을 주는 범위, 연결되어 있는 강도를 모두 반영한 값이다. 상대적 삼각매개중심성은 이 동시출현단어의 영향력을 주는 범위를 많이, 구조에 대한 위치를 적게 반영한 값이며, 평균연관성의 경우 연결되어 있는 강도를 많이 반영한 값이다. 노드 간의 간선의 경우 키워드 빈도가 클수록 굵게 표현되었으며, 상대적 삼각매개중심성 값에 따라 노드의 크기로 반영하였다. 평균연관성 상위 10개 노드들을 사각형으로 나타내었으며, 해당 키워드의 상대적 삼각매개중심성 지수 및 평균연관성 지수를 비교를 통해 건설 분야의 인공지능 적용 전반에 중심이 되는 키워드를 확인할 수 있었다. 표 3과 같이 상대적 삼각매개중심성과 평균연관성을 비교하였을 때, ‘deep learning(1)’, ‘artificial intelligence(2)’, ‘neural network(3)’, ‘machine learning(5)’, ‘deep learning algorithm(7)’, ‘construction industry(8)’, ‘previous studies(12)’, ‘learning algorithm(14)’는 모두 상위 10개 안에 포함되는 높은 전역중심성을 가지며, 이는 건설 분야의 인공지능 적용에 활용되는 키워드라고 할 수 있다. 또한 상대적 삼각매개중심성에서의

‘high accuracy(21)’, ‘recent years(25)’, 평균연관성에서의 ‘deep learning model(10)’, ‘input data(42)’ 키워드 역시도 건설 분야 인공지능 적용 연구에 자주 활용되는 키워드라고 볼 수 있다.

또한 건설 분야 인공지능 적용 연구에서 클러스터 간에 핵심 중개자 역할을 하는 키워드들을 확인하기 위하여 NodeXL을 이용하여 매개중심성 지수를 측정하였다. 이 매개중심성 값 상위 10위까지를 기준으로 deep learning(1), learning algorithm(14), compressive strength(19), input data(42), deep learning model(10), concrete structure(17), support vector machine(54), point clouds(13), deep learning algorithm(7), construction industry(8) 순이었다.

그림 4와 같이 노드의 크기는 지역성 중심에 따라, 간선의 크기는 키워드 간 빈도에 비례하도록 하며, 매개중심성 값이 4.3 이상인 상위 10개의 노드들을 삼각형 형태로 표시하여 네트워크 지도를 작성하였다.

다음으로 11개 클러스터의 세부 주제영역을 분석을 위하여 지역중심성 지수를 측정하였다. 또한 클러스터의 세부 키워드들을 이해하기 위해, 클러스터에서 영향력이 가장 높은 키워드를 확인할 수 있는 지역중심성의 지표인 상대적 최근접이웃중심성값을 이용하였다.

표 3. 중심성 상위 10개의 키워드 리스트
Table. 3. List of top 10 keywords for centrality

Rank	Keyword(no.)	Relative triangle betweenness centrality	Keyword(no.)	Average mean association
1	deep learning(1)	0.68723	deep learning(1)	0.14493
2	neural network(3)	0.627	neural network(3)	0.12028
3	previous studies(6)	0.55733	learning algorithm(12)	0.11023
4	deep learning algorithm(4)	0.54282	artificial intelligence(2)	0.10559
5	learning algorithm(12)	0.52322	deep learning algorithm(4)	0.1032
6	artificial intelligence(2)	0.48476	previous studies(6)	0.0925
7	machine learning(15)	0.45936	construction industry(5)	0.07069
8	recent years(22)	0.45428	deep learning model(10)	0.06894
9	high accuracy(18)	0.44557	input data(26)	0.06281
10	construction industry(5)	0.43759	machine learning(15)	0.05925

machine learning(5), learning algorithm(14)과 알고리즘 중 대표적인 random forest(30)와 support vector machine(55)로 구성된다. 3 클러스터는 5, 8, 11 클러스터와도 관련이 있는 것을 확인할 수 있다.

제 4 클러스터는 딥러닝 알고리즘(Deep learning algorithm)과 관련한 키워드 중 알고리즘이 사용되는 활용분야와 관련된 키워드로 구성되어 있다. 건설 지역, 이전 연구, 최신 기술, 장비, 관리, 종류, 성능 등의 construction site(9), previous studies(12), novel method(20), high accuracy(21), different types(27), feature extraction(29), important role(49), construction equipment(43), construction management(40), great significance(48), superior performance(51)로 매우 중요한 키워드들로 구성되어 있다.

제 5 클러스터는 건설 산업(Construction industry)과 관련된 키워드들로 구성되어 있다. construction industry(8), natural language processing(26), great potential(52)로 이루어져 있으며, 건설 산업과 관련하여 자연어 처리를 통한 연구가 진행되고 있으며, 인공지능을 적용하는 데에 있어서 잠재성을 가진 것을 확인할 수 있다. 2, 3, 8 클러스터와 연결되어 있다.

제 6 클러스터는 딥러닝 모델(Deep learning model)과 관련된 입력, 활용, 적용된 분야 관련한 concrete structure(17), long short-term memory(32), input data(42), automatic detection(50)로 구성되어 있으며, 1, 8 클러스터와 연결되어 있다.

제 7 클러스터는 건설 분야 관련 사람들을 나타내는 토목 공학자(Civil engineers)이며, 주 키워드는 civil engineers(23), construction worker(31)의 두 개의 키워드로 이루어져 ‘사람’ 관련 연구도 진행중임을 알 수 있다.

제 8 클러스터는 건설 분야의 자재 관련 키워드인 압축 강도(Compressive strength)의 주제로 compressive strength(19), accurate prediction(36), extreme gradient(46)로 구성되어 있으며, 3, 5, 6 클러스터와 연결되어 있다.

제 9 클러스터는 인공지능 기법 중 하나인 컴퓨터 비전(Computer vision)의 주제로 구성된 클러스터이며, 키워드로는 computer vision(6) crack detection(24), transfer learning(38)으로 구성되어 있으

며 이 컴퓨터 비전 기술은 크랙 감지와 전이 학습에 사용되는 키워드라는 것을 확실히 알 수 있다.

제 10 클러스터는 모델 성능(Model performance)와 관련된 construction project(15), model performance(28), prediction accuracy(37), large number(41), hybrid model(44)로 구성되어 있으며 성능 향상을 위해 연구되고 있는 요소와 모델 및 분야를 확인할 수 있었다.

마지막으로 제 11 클러스터는 인공지능에서 활용되는 알고리즘인 유전자 알고리즘(Genetic algorithm)과 관련한 키워드로 구성되어 있으며, genetic algorithm(39)와 이를 활용하는 sensitivity analysis(46) 키워드로 구성되어 있다. 또한 3, 10 클러스터와도 연결되어 있다.

IV. 클러스터 분석을 위한 지적구조

네트워크 분석을 활용한 클러스터 분석을 보완하기 위하여 SPSS ver 29.0을 통한 피어슨 상관계수를 활용한 2차 연관성 행렬로부터 클러스터 분석을 하였다. 클러스터링을 위해 Ward 기법과 덴드로그램을 표현하기 위하여 z점수 표준화 설정 및 제곱 유클리디안 거리를 이용하였다. 이를 통해 도출된 결과를 바탕으로 본 연구에서는 건설 분야에 인공지능 적용에 관련하여 가장 잘 나타내는 클러스터의 개수는 4개라 판단하여 표 4에 나타내었다. 클러스터명은 클러스터에 포함된 키워드들 중 클러스터를 나타내는 가장 적절한 키워드로 나타내었으며, 클러스터를 구성하는 키워드를 통해 건설 분야에 인공지능 적용에 관한 흐름을 파악할 수 있었다.

제 1 클러스터 deep learning은 건설 분야에 적용하고 있는 딥러닝에 관련한 키워드들로 구성되어 있다. deep learning model(10), point clouds(13), semantic segmentation(16), concrete structure(17), novel method(20), deep neural network(22), model performance(28), long short-term memory(32), fl score(35), training data(39), input data(42)을 포함하며 딥러닝의 입력, 출력, 성능, 평가, 활용 데이터 등과 같이 건설 분야에서 전반적으로 활용되어지는 딥러닝의 핵심 키워드들로 구성됨을 알 수 있다.

표 4. 클러스터 분석을 통한 지적 구조

Table. 4. Domain structure by cluster analysis

Cluster no.	Keyword(no.)	Cluster no.	Keyword(no.)
Cluster #1 (12 keywords) deep learning	deep learning(1) deep learning model(10) point clouds(13) semantic segmentation(16) concrete structure(17) novel method(20) deep neural network(22) model performance(28) long short-term memory(32) f1 score(35) training data(39) input data(42)	Cluster #3 (14 keywords) neural network	neural network(3) learning algorithm(14) construction project(15) compressive strength(19) random forest(21) accurate prediction(36) prediction accuracy(37) large number(41) hybrid model(44) genetic algorithm(45) extreme gradient(46) superior performance(51) sensitivity analysis(53) support vector machine(54)
Cluster #2 (13 keywords) artificial intelligence	artificial intelligence(2) machine learning(5) construction industry(8) building information modeling(11) civil engineers(23) construction worker(31) future research(18) recent years(25) natural language processing(26) big data(33) systematic review(47) great significance(48) great potential(52)	Cluster #4 (15 keywords) deep learning algorithm	deep learning algorithm(4) convolution neural network(4) computer vision(6) construction site(9) previous studies(12) high accuracy(21) different types(27) mean average precision(34) crack detection(24) feature extraction(29) transfer learning(38) construction management(40), construction equipment(43) important role(49) automatic detection(50)

제 2 클러스터는 artificial intelligence(2)가 대표 키워드로써, 인공지능과 관련된 키워드들로 구성되어 있다. 최근 건설 산업에서도 인공지능이 적용되는 사례가 많아 이에 대표되는 키워드인 인공지능의 대표되는 키워드인 machine learning(5), construction industry(8), 전통적인 인공지능 자동화 building information modeling(11), 건설 분야에 종사하는 사람들 civil engineers(23), construction worker(31)와 연구에 대한 future research(18), recent years(25), 제일 많이 쓰는 인공지능 기법 중 하나인 natural language processing(26), 활용분야인 big data(33), 체계적인 리뷰인 systematic review(47), 건설 산업에 대한 중요성을 나타내는 great significance(48), great potential(52) 키워드로 구성되어 진다.

neural network(3)로 대표되는 제 3 클러스터는 건설 분야에서 활용되는 인공지능의 성능을 높이기 위

한 알고리즘에 대한 클러스터이며, 학습 알고리즘 learning algorithm(14), 압축 강도 compressive strength(19), 인공지능 기법 중 하나인 random forest(21), hybrid model(44), genetic algorithm(45), extreme gradient(46)와 support vector machine(54) 성능에 관한 superior performance(51), 해당 알고리즘의 정확도에 관한 accurate prediction(36), prediction accuracy(37), large number(41), 적용분야를 나타내는 민감성 분석 sensitivity analysis(53), 이러한 인공지능 활용을 통해 성능향상을 construction project(15) 연구를 통해 이루어짐을 알 수 있다.

마지막 제 4 클러스터는 deep learning algorithm(7)으로 대표되는 키워드로 딥러닝 알고리즘의 지난 연구와 적용, 활용되는 키워드로 구성되어 있다. 대표적인 이미지 분석을 위한 딥러닝 알고리즘인 convolution neural network(4), computer vision(6), 딥

VI. 결 론

본 연구는 2018년부터 2022년까지 최근 5년간 Web of Science에서 건설 분야에서의 인공지능 적용을 주제로 하는 문헌 데이터들을 수집하고 동시출현단어 행렬을 활용한 네트워크 분석, 클러스터 분석, 다차원축적지도를 통해 건설 분야에 인공지능 적용 연구에 대한 지적구조를 살펴보았다. 네트워크 분석을 위해 유사도 및 피어슨 상관계수를 통한 연관성 행렬을 생성하고 이들의 관계를 패스파인더 네트워크 알고리즘을 통한 시각화 및 최근접 이웃 클러스터링 알고리즘을 통한 클러스터 분석을 통해 건설 분야 인공지능 적용에 관한 세부 주제영역을 확인하였다.

전역 및 지역 중심성 분석을 통해 건설 분야 인공지능 적용의 중심이 되는 키워드와 클러스터에서의 영향력이 높거나 클러스터 간에 매개가 되는 키워드들을 분석하였다. 네트워크 분석을 보완하기 위한 클러스터 분석을 수행하고, 이에 대한 결과를 다차원축적지도에 반영하여 건설 분야 인공지능 적용에 관한 지적구조와 세부 주제영역을 분석하였다. 네트워크 분석의 결과, 11개의 클러스터가, 클러스터 분석의 결과로 4개의 클러스터가 도출되었다. 이를 통해 건설 분야에 인공지능 적용 연구의 흐름이 크게는 건설 분야의 deep learning 적용 방안, artificial intelligence에 관한 연구와 분석, neural network를 통한 정확도 향상, deep learning algorithm을 통한 건설 분야의 적용에 관한 연구들이 수행되고 있고, 세부적으로는 deep learning, artificial intelligence, neural network, deep learning algorithm, construction industry, civil engineers, computer vision, model performance, genetic algorithm 주제로 연구되고 있음을 알 수 있었다.

이상의 결과를 종합해 보면 2018년부터 2022년까지의 공학 토목과 건설 건축 기술 범주에서 건설 분야 인공지능 적용에 관한 핵심적인 연구 영역은 딥러닝 활용, 건설 산업, 성능 향상, 기존 연구 분석이다. 이 결과는 건설 분야에 인공지능의 ‘적용’이 아직 초기 단계에 머물러 있고, 보다 다양하고 심층적인 분석 및 연구가 필요하다는 점을 보여주고 있다.

본 연구는 대상 분야가 토목 공학 및 건설, 건축 기술에 국한되었기 때문에 분석 결과가 모든 건설 분야에 적용될 수 있는 것은 아니다. 그러나 본 연구는 건설 분야에서도 인공지능 적용에 관한 지적구조를 분석할 수 있음을 보여주는 것에 큰 의의가 있으며, 향후 지속적인 연구에 도움이 될 것으로 기대한다.

References

- [1] P. Ongsulee, "Artificial intelligence, machine learning and deep learning", 2017 15th International Conference on ICT and Knowledge Engineering, Bangkok, Thailand, pp. 1-6, Nov. 2017. <https://doi.org/10.1109/ICTKE.2017.8259629>.
- [2] K. G. Kim, "Book Review: Deep Learning", Healthcare Informatics Research, Vol. 22, No. 4, pp. 351-354, Oct. 2016. <https://doi.org/10.4258/hir.2016.22.4.351>.
- [3] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: methods and applications", Journal of Manufacturing Systems, Vol. 48, No. C, pp. 144-156, Jul. 2018. <https://doi.org/10.1016/j.jmsy.2018.01.003>.
- [4] R. Vargas, A. Mosavi, and R. Ruiz, "Deep Learning: A Review", Advances in Intelligent Systems and computing, preprints, pp. 1-10, Oct. 2018. <https://doi.org/10.20944/preprints201810.0218.v1>.
- [5] S. K. Seo and E. K. Chung, "Domain Analysis on the Field of Open Access by Co-Word Analysis", Journal of the Korean bible society for library and information science, Vol. 24, No. 1, pp. 207-228, Mar. 2013. <https://doi.org/10.14699/kbiblia.2013.24.1.207>.
- [6] J.-Y. Lee, "A Study on the Network Generation Methods for Examining the Intellectual Structure of Knowledge Domains", Journal of the Korean Library and Information Science Society, Vol. 40, No. 2, pp. 333-355, Jun. 2006. <https://doi.org/10.4275/KSLIS.2006.40.2.333>.

- [7] J.-Y. Lee, "A Comparison Study on the Weighted Network Centrality Measures of tnet and WNET", Journal of the Korean Society for Information Management, Vol. 30, No. 4, pp. 241-264, Dec. 2013. <https://doi.org/10.3743/KOSIM.2013.30.4.241>.
- [8] D. Hansen, B. Shneiderman, and M. A. Smith, "Analyzing social media networks with NodeXL: Insights from a connected world", Journal of Human-Computer Interaction, Vol. 27, No. 4, Feb. 2011. <https://doi.org/10.1080/10447318.2011.544971>.
- [9] H. D. White and B. C. Griffith, "Author cocitation: A literature measure of intellectual structure", Journal of the American Society for Information Science, Vol. 32, No. 3, pp. 163-171, May 1981. <https://doi.org/10.1002/asi.4630320302>.
- [10] J.-Y. Lee, "A Study on the Network Generation Methods for Examining the Intellectual Structure of Knowledge Domains", Journal of the Korean Library and Information Science, Vol. 40, No. 2, pp. 333-355, Jun. 2006. <https://doi.org/10.4275/KSLIS.2006.40.2.333>.
- [11] J.-Y. Lee, "A Novel Clustering Method for Examining and Analyzing the Intellectual Structure of a Scholarly Field", Journal of the Korean Society for Information Management, Vol. 23, No. 4, pp. 215-231, Nov. 2006. <https://doi.org/10.3743/KOSIM.2006.23.4.215>.
- [12] J.-Y. Lee, "Centrality Measures for Bibliometric Network Analysis", Journal of the Korean Library and Information Science, Vol. 40, No. 3, pp. 191-214, Sep. 2006. <https://doi.org/10.4275/KSLIS.2006.40.3.191>.

저자소개

김 선 검 (Sun-Kyum Kim)



2010년 2월 : 세종대학교
컴퓨터공학과(공학사)
2016년 2월 : 연세대학교
컴퓨터과학과(공학박사)
2019년 3월 :
한국과학기술정보연구원
2020년 7월 :

차세대융합기술연구원

2020년 8월 ~ 현재 : 한국건설기술 연구원

미래스마트건설연구본부 수석연구원

관심분야 : 인공지능, 데이터분석, 블록체인

원 지 선 (Jisun Won)



2003년 2월 : 경희대학교
토목건축공학부 (공학사)
2005년 2월 : 경희대학교
건축공학과(공학석사)
2022년 3월 ~ 현재 : 경희대학교
건축학과 박사과정

2005년 12월 ~ 현재 : 한국건설기술

연구원 미래스마트건설연구본부 수석연구원

관심분야 : 데이터모델링(STEP, IFC) BIM(Building Information Modeling), 인공지능, 자연어처리

신 재 영 (Jaeyoung Shin)



2015년 2월 : 한양대학교
실내건축디자인학과(이학사)
2017년 2월 : 한양대학교
실내건축디자인학과(이학석사)
2020년 9월 ~ 현재 : 연세대학교
실내건축학과 박사과정

2017년 3월 ~ 현재 : 한국건설기술

연구원 미래스마트건설연구본부 전임연구원

관심분야 : 실내건축(설계, 실내동선), BIM(Building Information Modeling), 인공지능