

건설산업 내 기계학습 알고리즘 (Machine Learning Algorithm) 활용 동향 (Machine Learning Algorithms For Construction Industry)

KICEM



이광표 한국건설산업연구원 부연구위원

I. 서론

과거 건설산업에서의 데이터 활용도는 낮은 수준이었으며, 의사결정 시 전문가의 경험 또는 직관에 의존하여 왔다. 그러나, 최근 들어 건설산업 내 존재하는 다양한 비정형 데이터 (Unstructured Data)를 정형화·구조화(Structuration)할 수 있게 되고, 분석목적에 적합한 데이터를 적극적으로 수집할 수 있는 다양한 기술들이 도입되면서 데이터의 활용도가 점진적으로 증가하고 있다. 또한, 건설산업 내 존재하는 이러한 빅데이터 (Big Data)를 분석하기 위한 데이터 마이닝(Data Mining), 인공지능(Artificial Intelligence) 알고리즘 등의 다양한 기술들이 도입 및 적용되고 있는 상황이다. 특히, 2016년 알파고의 등장으로 주목받기 시작한 기계학습 또는 머신러닝 알고리즘(Machine Learning Algorithm)은 빅데이터를 분석하기 위한 새로운 접근법으로 급속도로 발전하고 있다. 이미 정보통신, 교육, 의료, 제조, 금융 등의 사회 각 분야에서는 다양한 기계학습 알고리즘을 활용한 데이터 분석을 수행하고 있으며, 새로운 상황 또는 조건에 대한 미래 결과를 예측함으로써 보다 높은 수준의 서비스를 제공하고자 노력하고 있다. 현재 건설산업에서도 이러한 기계학습 알고리즘들에 대한 적용성 파악을 위해 건설 생애주기(Life-Cycle) 동안 발생하는 다양한 데이터를 활용하여, 건축물 에너지관리, 공사비 및 공사기간 성과예측, 센싱(Sensing) 데이터 분석을 통한 새로운 지식(Finding) 창출 등의 다양한 분야에서 연구를 시작하고 있는 시점이다 (Cheng et al, 2010; Huang et al, 2012; Naganathan et al, 2016; Wauters and Vanhoucke 2016). 이러한 배경하에, 본 고에서는 기계학습 알고리즘들에 대한 기본적인 개념 및 분류에 대해 간략히 설명하고, 건설산업 내 활용현황 및 최신 연구동향에 대해 소개하고자 한다.

II. 기계학습 유형 및 관련 알고리즘

기계학습은 컴퓨터 스스로가 방대한 양의 데이터를 분석해 미래를 예측하는 기술을 말한다. 구체적으로, 특정한 목적에 적합한 입력 데이터를 트레이닝(Training)하고 검증(Validation)함으로써 새로운 상황에 대한 예측(Prediction or Estimation), 분류(Classification), 또는 그룹화(Grouping or Clustering)하는 것을 의미한다. 이러한 머신러닝은 학습되는 데이터의 종류 및 형태에 따라 일반적으로 지도학습(Supervised Learning), 비지도학습(Unsupervised Learning), 준지도학습(Semi-Supervised Learning)으로 크게 분류할 수 있으며, 이 외에도 강화학습(Reinforcement Learning), 심화학습(Deep Learning) 등이 있다. 또한, 각각의 학습방법에 해당하는 다양한 알고리즘들이 꾸준히 개발되고 있는 상황이다.

- 지도학습이란, 독립변수에 대한 종속변수(Numerical Value or label)가 존재하는 데이터를 활용한 학습방법으로, 데이터의 구조를 트레이닝함으로써 모델을 제안하고 새로운 독립변수에 대한 결과 값을 예측한다.
- 반면, 비지도학습은 지도학습과는 달리 종속변수에 대한 값이 주어지지 않아 독립변수에 따른 종속변수의 성격이 정의되지 않은 데이터를 트레이닝하여 그룹 또는 무리(Group or Cluster)를 도출하는 학습방법이다. 이 방법은 일반적으로 데이터들의 분포를 활용한다.
- 마지막으로, 준지도학습은 지도학습과 비지도학습의 속성을 동시에 가지는 학습방법으로 종속변수에 대한 값을 가지는 데이터와 가지지 않는 데이터를 모두 트레이닝시킴으로써 새로운 독립변수에 대한 예측 정확

도를 증가시킨다는 장점을 가지고 있다.

위와 같은 학습방법들은 분석목적에 따라 예측 또는 추정 (Prediction or Estimation), 분류 (Classification), 패턴 (Pattern)이나 구조의 파악, 그룹화 (Grouping), 차원 축소 (Dimension Reduction), 문자, 영상, 이미지 (Text, Video, Image) 등에 대한 분석기능을 수행한다. 각 기능을 대표하는 알고리즘으로는 예측 및 추정의 경우, 인공신경망 (Artificial Neural Network, ANN), SVM (Support Vector Machine) 등이 있고, 분류 목적의 알고리즘으로는 k-NN (k-Nearest Neighbor), 로지스틱 회귀(Logistic Regression), Ensemble (Bagging, Boosting, Random Forest) 기법, SVM, 의사결정트리(Decision Tree) 등이 있다. 이 외에도 K-평균 군집화 (K-means Clustering)과 같은 그룹화 알고리즘과 주성분분석 (Principle Component Analysis, PCA) 등의 차원을 축소하기 위한 알고리즘들이 존재한다.

표1. 기계학습 유형 및 대표 알고리즘

기계학습 종류	기능 (Function)	관련 알고리즘
지도학습 (Supervised Learning)	분류 및 추정	Artificial Neural Network(ANN) Support Vector Machine(SVM) Logistic Regression Ensemble(Bagging, Boosting, Random Forest) Decision Tree 외 다수 알고리즘 존재
비지도학습 (Unsupervised Learning)	그룹화, 차원축소 및 이상탐지	K-means Clustering Principle Component Analysis(PCA) Anomaly Detection

III. 글로벌 연구 동향

기계학습 분야가 주목받지 못했던 1990년대 또는 2000년대 초기까지는 건설분야에 기계학습 알고리즘을 적용한 사례가 비교적 적은 수준이었으며, 데이터 중심의 연구를 주장하는 문헌들에서 간간히 언급되고 있었다. 그러나, 2000년대 중반 이후, 빅데이터에 대한 개념이 제시되기 시작하고 이를 처리하기 위한 다양한 인공지능 및 데이터마이닝 알고리즘들이 등장하면서 본격적으로 기계학습 알고리즘을 활용한 연구들이 시작되었다. 현재는 건설산업 내 다양한 분야에서 머신러닝 알고리즘의 적용성을 파악하기 위한 연구들이 수행되고 있다. 건설산업 내 머신러닝 알고리즘이 적용되고 있는 대표적인 분야로는 건설사업관리, 건축물 유지관리, 에너지 분석, 센싱 데이터 및 컴퓨터 영상처리

분야 등이 있으며, 연구 초기단계로 각 분야에의 활용가능성 파악에 중점을 두고 연구가 수행되고 있다.

(1) 건설사업관리 분야

건설산업 내 기계학습 알고리즘이 도입되고 있는 대표적인 분야로는 다양한 건설사업관리 분야를 들 수 있다. 예를 들면, 정형 또는 비정형 데이터들이 다수 존재하는 안전관리분야에의 적용성 파악이 이루어지고 있으며, 과거 사업수행사례를 중심으로 신규사업의 공사비 및 공사기간에 대한 성과예측을 위한 연구들도 수행되고 있었다. 이 외에도 사업 타당성을 분석하거나 계약자를 합리적으로 선정하는 과정에서 머신러닝을 활용한 연구들이 존재하였다.

먼저, Tixier et al. (2016)는 안전사고 사례보고서(Safety Report)를 분석하여, 사고를 발생시키는 속성(Attributes)들을 정의하고, 정의한 속성들에 대한 데이터를 활용하여 새로운 상황에서 발생이 예상되는 안전사고를 예측하였다. 이 과정에서 데이터 분석 및 예측의 정확성 증진을 위해 랜덤포레스트와 Stochastic Gradient Tree Boosting (SGTB)를 적용하여, 사고를 발생시키게 되는 에너지 원인(Energy Source), 부상 유형 (Injury Type), 부상 부위(Body Part), 부상 정도(Injury Severity)를 예측해냈다. Cheng et al. (2010)는 사업종료 시점의 공사비(Estimate at Completion, EAC)를 예측하기 위해 fast messy Genetic Algorithm (fmGA)과 SVM을 적용한 Evolutionary SVM Inference Model(ESIM)을 제안하고, 이를 과거 건설사업 데이터에 적용해봄으로써 제안한 모델의 적용성을 파악하였다. Wauters and Vanhoucke (2016)는 건설사업의 공사기간 예측을 위해 Bagging과 Boosting 기법을 적용한 의사결정트리를 제안 및 적용하고, 적용결과를 SVM알고리즘과 성능을 비교해봄으로써 제안한 모델의 효용성을 파악하였다. Asadi et al. (2015) 역시 사업초기단계에 공사기간 지연여부 및 기간을 예측하기 위해 의사결정트리를 적용한 바 있다.

이 외에도 Lam et al. (2009)의 경우, 합리적인 계약자 선정을 위해 SVM 알고리즘을 활용하여 입찰참가자격 사전심사 (Pre-qualification)를 수행하고, 이를 인공신경망 알고리즘과 비교함으로써 SVM 알고리즘의 성능을 파악하였다. Skibniewski et al. (1997)는 사업초기 건설사업 타당성 분석을 위한 요인들을 도출해내고, 각 요인들에 대한 정성적 평가결과를 활용하여 신규사업의 타당성을 예측하였다.

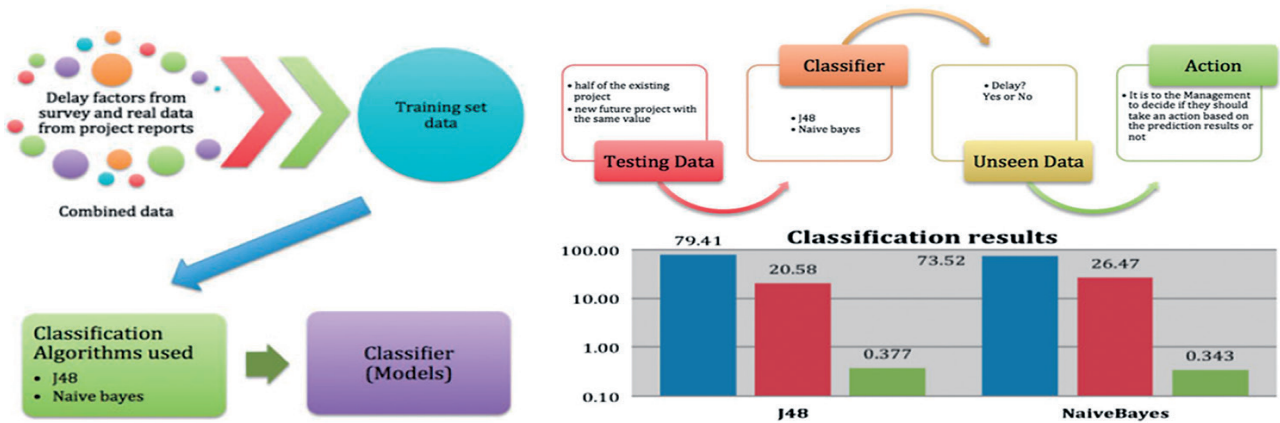


그림 1. 지도학습 (Supervised Learning) 적용 예시 (Asadi et al. 2015)

(2) 건축물 유지관리 및 에너지 분석

건축물 유지관리 분야와 관련하여 Catbas and Malekzadeh (2016)는 가동교(Movable Bridge)의 손상탐지(Damage Detection)를 목적으로 기계부품에 가속도계(Accelerometer) 센서를 적용함으로써 데이터를 수집하고 머신러닝 알고리즘을 적용하였다. Kargah-Ostadi (2014)는 포장유지관리 시스템 (Pavement Management System) 구성에 있어 중요한 요소인 포장성능(Pavement Performance) 예측을 위해 인공신경망, 방사기초함수(Radial Basis Function, RBF) 네트워크, 그리고 SVM 알고리즘을 적용하여 각 알고리즘 간의 성능을 비교분석 하였다. 그리고 건축물 에너지 분석과 관련해서, Naganathan et al. (2016)는 건축물 및 변전소의 실시간 데이터(Real-Time Data)에 대해 군집화 알고리즘(Clustering Algorithm)을 적용하여 건축물의 에너지 수요-공급에 대한 분석을 수행한 바 있다. 또한, 이 과정에서 준지도학습 알고리즘을 적용하여 건축물 에너지 손실을 최적화하고 손실요인에 대한 파악을 수행하였다. Tsanas and Xifara (2012) 역시 주거 건축물의 정확한 에너

지 성능예측을 위해 랜덤포레스트 알고리즘을 적용하고, 이를 선형회귀(Linear Regression)방법과 비교함으로써 랜덤포레스트 알고리즘의 에너지 분석에 대한 적용가능성을 파악하였다.

(3) 센싱 데이터 및 컴퓨터 영상처리 (Video and Image processing)

앞서 소개한 적용분야 외에도 다양한 종류의 센서 및 영상으로부터 얻은 대용량 데이터를 활용하여 기계학습 알고리즘의 건설산업 내 활용분야를 파악하고 있다. 예를 들어, Huang et al. (2012)는 머신러닝을 활용하여 화재 등의 건물 재난을 사전에 예측하기 위해, 무선 네트워크 기술 및 온도센서로부터 얻을 수 있는 실시간 데이터를 분석하였고, Yang et al. (2010)는 비디오 영상데이터를 분석하여 건설현장 내 다중작업자의 위치를 추적하기 위한 방안을 제안하였다. 또한, Son et al. (2011)은 영상 속 존재하는 색상 데이터를 분석하여 콘크리트 구조물을 자동으로 감지해 낼 수 있는 방법을 제안하였다.

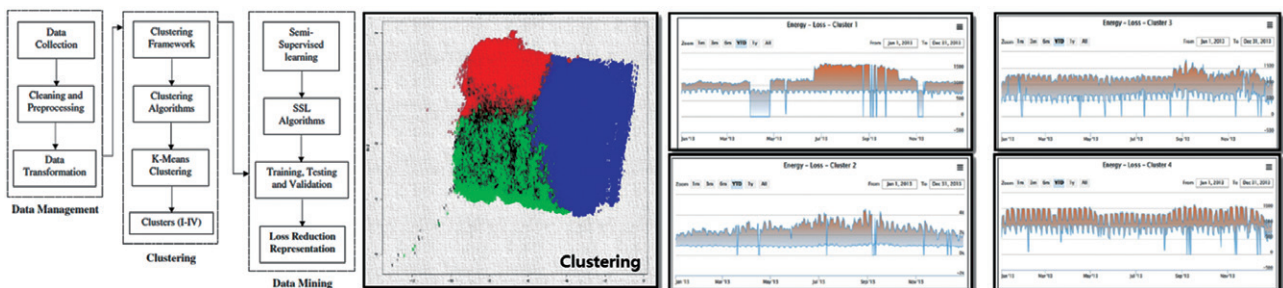


그림 2. 비지도학습 (Unsupervised Learning) 및 준지도 학습 (Semi-Supervised Learning) 적용 예시 (Naganathan et al. 2016)

IV. 결론

본 고에서 간략하게 소개한 바와 같이 현재 다양한 기계학습 알고리즘들이 개발된 상태이며, 건설산업에서는 이에 대한 분야별 적용성을 파악하기 위한 연구가 이루어지고 있었다. 이러한 상황에서, 기계학습 알고리즘을 적극적으로 실무에 적용하기 위해서는 몇 가지 기술적 한계점들에 대한 극복이 필요할 것으로 사료된다. 우선, 머신러닝은 컴퓨터 스스로가 대용량의 데이터를 학습하고 처리하는 기술이기 때문에, 기존 건설산업 내에 존재하는 활용가능한 데이터들에 대한 확보가 필수적일 것이다. 이를 위해서는, 정량데이터뿐 아니라 보고서 또는 문서의 형태로 존재하는 데이터에 대한 면밀한 파악이 요구될 것이며, 이러한 데이터를 직접적으로 활용하기 위한 사전처리(Pre-processing)가 필요할 것이다. 예를 들면, 텍스트 마이닝(Text Mining) 기법을 통해 비정형 데이터로 존재하는 각종 문서를 정형화시킴으로써 데이터베이스(Database)를 구축할 수 있을 것이다. 한편, 이러한 기술적 한계점은 현재의 기술발전 수준을 고려한다면 머지 않은 미래에 극복가능할 것으로 예상하며, 발전된 머신러닝 기술은 건설산업 내 존재하는 불확실성을 파악할 수 있는 중요한 미래예측 방법으로 활용될 것이라 전망한다. 또한, 이는 건설기업의 운영 및 사업 수행 시 계획의 정확도를 증진시키고, 잠재되어 있는 다양한 건설지식 발견에 기여할 것이다.

참고문헌

- Asadi, A., Alsubaey, M., and Makatsoris, C. (2015). "A machine learning approach for predicting delays in construction logistics." *International Journal of Advanced Logistics*, 4(2), 115–130.
- Catbas, F. N., and Malekzadeh, M. (2016). "A machine learning-based algorithm for processing massive data collected from the mechanical components of movable bridges." *Automation in Construction*, 72, 269–278.
- Cheng, M. Y., Wu, Y. W., and Wu, C. F. (2010). "Project success prediction using an evolutionary support vector machine inference model." *Automation in Construction*, 19(3), 302–307.
- Elazouni, A. M. (2006). "Classifying construction contractors using unsupervised-learning neural networks." *Journal of Construction Engineering and Management*, 132(12), 1242–1253.
- Huang, Q., Cox, R. F., Shaurette, M., and Wang, J. (2012). "Intelligent building hazard detection using wireless sensor network and machine learning techniques." In *Computing in Civil Engineering*, 485–492.
- Kargah-Ostadi, N. (2014). "Comparison of machine learning techniques for developing performance prediction models." In *Computing in Civil and Building Engineering*, 1222–1229.
- Lam, K. C., Palaneeswaran, E., and Yu, C. Y. (2009). "A support vector machine model for contractor prequalification." *Automation in Construction*, 18(3), 321–329.
- Naganathan, H., Chong, W. O., and Chen, X. (2016). "Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches." *Automation in Construction*, 72, 187–194.
- Skibniewski, M., Arciszewski, T., and Lueprasert, K. (1997). "Constructability analysis: machine learning approach." *Journal of computing in civil engineering*, 11(1), 8–16.
- Son, H., Kim, C., and Kim, C. (2011). "Automated color model-based concrete detection in construction-site images by using machine learning algorithms." *Journal of Computing in Civil Engineering*, 26(3), 421–433.
- Tsanas, A., and Xifara, A. (2012). "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools." *Energy and Buildings*, 49, 560–567.
- Tixier, A. J. P., Hollowell, M. R., Rajagopalan, B., and Bowman, D. (2016). "Application of machine learning to construction injury prediction." *Automation in construction*, 69, 102–114.
- Wauters, M., and Vanhoucke, M. (2016). "A comparative study of Artificial Intelligence methods for project duration forecasting." *Expert Systems with Applications*, 46, 249–261.
- Yang, J., Arif, O., Vela, P. A., Teizer, J., and Shi, Z. (2010). "Tracking multiple workers on construction sites using video cameras." *Advanced Engineering Informatics*, 24(4), 428–434.

■ 이광표 E-mail: leekp@cerik.re.kr