

머신러닝 회귀분석 유형에 따른 총 공사기간 예측 모델에 관한 연구

A Study on the Prediction Model of Total Construction Period according to the Type of Machine Learning Regression

강윤호¹ · 윤석현^{2*}

Kang, Yun-Ho¹ · Yun, Seok-Heon^{2*}

Abstract : In construction work, there is often a difference between the estimated construction period and the actual construction period. Accordingly, the project may be delayed from the scheduled date, leading to huge losses due to problems such as increased costs during construction. In this way, it is important to calculate the appropriate construction period at the project planning stage in construction work. To solve this problem, we would like to study a model that will increase the accuracy of the scheduled construction period at the project planning stage. This study compared and analyzed linear regression, Lasso regression, Ridge regression among the types of regression analysis to select an appropriate construction period prediction model to secure an appropriate construction period at the project planning stage to reduce problems during construction.

키워드 : 공사기간, 머신러닝, 회귀분석

Keywords : construction period, machine learning, regression analysis

1. 서론

1.1 연구의 배경 및 목적

건축 프로젝트는 환경, 목적 등 복잡하고 복합적인 특성에 의해 합리적인 공사계획이 중요하다. 특히, 건설공사에 산정된 예정 공사기간과 실제 공사기간 사이에 많은 차이가 발생한다. 이에, 프로젝트가 예정 공사기간보다 지연되는 경우가 발생하여 시공 중 비용증가, 시민 불편증가 등의 문제점으로 막대한 손실로 이어진다. 이와 더불어 설계단계에서 공사 기간을 무리하게 산정하여 공기 부족, 안전사고 등의 문제점이 발생한다. 문제점을 줄이기 위해 2019년에는 국토부에서 공공 건설공사의 공사기간 산정기준 제정안을 훈령으로 발표했다[1]. 이러한 노력에도 불구하고 아직까지 적정공사의 산정보다는 공사비의 절감에 더 많은 관심을 가지고 있으며 관련 연구가 부족한 실정이다[2]. 따라서 공사기간 산정에 있어 추가적인 연구가 필요하고 공사기간 산정의 정확도를 높여야 할 필요가 있다.

본 연구는 공사기간 산정 모델의 정확도를 높이기 위해 다양한 회귀분석 유형을 적용하여 공사기간 예측 모델을 구축하고자 한다. 모델에 적용한 회귀분석 유형으로 선형회귀, 라쏘회귀, 릿지회귀를 사용하여 비교·분석한다. 다양한 회귀모델 중 공사기간 예측 정확도를 향상시키는 최적의 예측 모델을 제시하고자 한다.

2. 공사기간 산정모델

2.1 공사기간 산정모델 구성

본 연구는 머신러닝 회귀분석 유형에 따른 총 공사기간 예측 모델을 구축하기 위해 국토교통부의 적정 공사기간 확보를 위한 가이드라인, 서울특별시의 공정관리 지침을 참고하여 영향요인을 선정하였다. 영향요인은 연면적, 건축면적, 최고 높이, 최대 지상 층수, 최대 지하 층수, 대지면적, 조정면적, 기준층 층고, 주차대수로 총 9가지로 정의한다.

예측 모델을 정의할 때 사용된 데이터는 2017년부터 2022년까지 조달청에서 5년간 발주한 789개 데이터를 사용한다. 이 중 이상치 확인 및 제거 프로세스를 통해 총 61건의 이상치가 제거되었다. 모델에 사용되는 최적화 방식, 노드 수, 학습 횟수, 옵티마이저 등은 동일하게 설정하였고 Visual Studio Code 개발 환경을 바탕으로 Keras 라이브러리로 모델의 성능을 비교·분석하였다. 표 1은 모델 구조를 나타낸 표이다.

1) 경상국립대학교, 석사과정

2) 경상국립대학교, 교수, 교신저자(gyfun@gnu.ac.kr)

표 1. 모델 구조

모델	회귀분석 유형	옵티마이저	활성화 함수	노드	학습횟수
Model 1	선형 회귀	Adam	ELU	6, 100, 64, 32, 16, 1	500
Model 2	라쏘 회귀	Adam	ELU	6, 100, 64, 32, 16, 1	500
Model 3	릿지 회귀	Adam	ELU	6, 100, 64, 32, 16, 1	500

2.2 공사기간 산정모델 결과

그림 1부터 그림 3은 본 연구의 회귀분석 유형에 따른 총 공사기간 예측 모델에서 학습 횟수에 따른 Loss 변화를 나타낸 모델의 결과 그래프이다. 이를 통해 학습데이터와 검증데이터의 수렴점과 차이점을 확인할 수 있어 모델의 과대 적합, 과소 적합을 파악할 수 있다. 그림 1은 학습이 진행됨에 따라 학습데이터는 0에 수렴하지만 검증데이터는 0에 수렴하지 않는다. 이는 본 모델에서 적합한 모델이 아닐 것으로 판단된다. 그림 2는 학습이 진행됨에 따라 학습데이터와 검증데이터의 Loss 값 전체 양상이 0에 수렴한다. 이는 학습 횟수가 500 이상일 때 더 최적의 값을 찾을 수 있을 것으로 판단된다. 그림 3은 학습이 진행됨에 따라 학습데이터와 검증데이터의 Loss 값 전체 양상이 0에 수렴하지만 학습 횟수 300 이상부터는 두 개의 데이터 Loss 값이 점점 차이가 벌어지는 것을 볼 수 있다. 이를 통해 그림 2의 그래프는 학습 횟수가 300에서 최적의 값을 찾을 수 있을 것으로 판단된다.

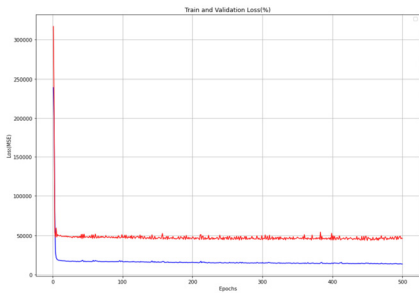


그림 1. Model 1 그래프

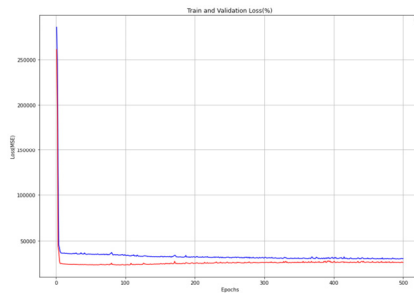


그림 2. Model 2 그래프

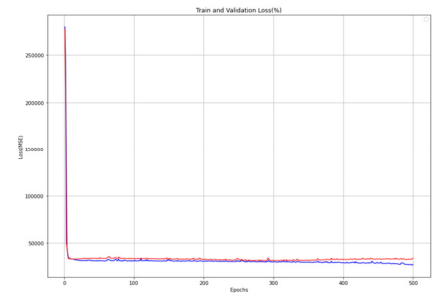


그림 3. Model 3 그래프

3. 결론

본 연구는 회귀분석 유형에 따른 총 공사기간 예측 모델을 비교하여 최적의 모델을 찾고자 하였다. 회귀분석은 하나의 종속변수가 다른 독립변수들에 의해 어떻게 설명 또는 예측되는지를 알아보기 위해 적절한 함수로 표현하여 자료분석을 하는 통계적인 기법이다 [3]. 회귀 유형 중 선형회귀, 라쏘회귀, 릿지회귀를 모델에 사용한다. 총 3개의 모델 중 모델 3 릿지 회귀를 사용한 모델은 학습데이터와 검증데이터의 Loss 값의 전체 양상이 비슷한 것으로 보아 과대적합 위험성이 발생할 것이라 판단되고, 모델 1은 데이터 간의 차이가 많이 발생하는 것으로 보아 과소적합이 발생할 것이라 판단한다. 이에 비해 라쏘 회귀를 사용한 모델이 다른 회귀를 사용한 모델보다 그래프 양상이 좋고 데이터 간의 차이가 많지 않아 모델 성능이 좋을 것이라 판단된다. 이에 본 연구에서 적합한 회귀모델은 라쏘회귀가 적합하다고 판단된다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 2023년도 지원으로 수행되었음(과제번호: RS-2021-KA163269).

참고문헌

1. 김승우. DNN을 통한 골조공사기간 예측에 관한 연구. 한양대학교 공학대학원. 2021. p. 48.
2. 김재준. 적정공기 판정. 한국설비기술협회지. 2006. p. 40.
3. 조용현. 신경망을 이용한 고신뢰성의 회귀분석 모델. 정보처리학회논문지. 2001. pp. 327-334.