

http://dx.doi.org/10.17703/JCCT.2023.9.1.685

JCCT 2023-1-84

머신러닝 기반의 실시간 자동화계측 데이터 분석 기법 연구

A Study on Machine Learning-Based Real-Time Automated Measurement Data Analysis Techniques

최정열*, 한재민**, 안대희***, 정지승****, 김정호*****, 이성진*****

Jung-Youl Choi*, Jae-Min Han**, Dae-Hui Ahn***, Jee-Seung Chung*****,
Jung-Ho Kim*****, Sung-Jin Lee*****

요약 도시의 인구증가 및 고밀화에 따라 기존 지하구조물에 인접하여 도심도 굴착 공사 물량이 증가하는 추세인 것으로 분석되었다. 현재 지하구조물 및 궤도는 외부요인에 의해 지하구조물의 손상이 다수 발생하는 실정이며 터널 내의 계측결과로 원인을 분석하여 예방차원이 아닌 사후처리에 대해서 측정을 하고 있는 실정이다. 본 연구의 목적은 공용중인 도시철도 선로와 인접한 굴착공사에 따른 구조물의 변형에 미치는 영향을 분석하는 것이다. 또한 외적 요인으로 인해 지하구조물 및 궤도 손상 및 파괴가 발생하기 전 구조물의 변위를 머신러닝 기법을 통해 구조물의 안전성을 평가하고자 한다. 분석결과, 분석한 데이터셋에서 구조물관리기준치에 도달하는 시간을 예측하기에 적합한 모델은 다항회귀 머신러닝 알고리즘인 것으로 분석되었다. 그러나 본 연구에서 적용한 자동화계측 데이터에 한정될 수 있으므로 추가적으로 구조물 조건의 다양성과 데이터양을 늘리는 향후 연구가 필요하다.

주요어 : 인접굴착, 지하구조물, 머신러닝, 실시간, 시간예측

Abstract It was analyzed that the volume of deep excavation works adjacent to existing underground structures is increasing according to the population growth and density of cities. Currently, many underground structures and tracks are damaged by external factors, and the cause is analyzed based on the measurement results in the tunnel, and measurements are being made for post-processing, not for prevention. The purpose of this study is to analyze the effect on the deformation of the structure due to the excavation work adjacent to the urban railway track in use. In addition, the safety of structures is evaluated through machine learning techniques for displacement of structures before damage and destruction of underground structures and tracks due to external factors. As a result of the analysis, it was analyzed that the model suitable for predicting the structure management standard value time in the analyzed dataset was a polynomial regression machine. Since it may be limited to the data applied in this study, future research is needed to increase the diversity of structural conditions and the amount of data.

Key words : Adjacent excavation, Underground structure, Machine learning, Real-time, Time prediction

1. 서 론

현재 도시철도는 내, 외부요인으로 인한 지하구조물 및 궤도의 안전성 모니터링 수요 급증, 도시의 인구증가

*정희원, 동양대학교 건설공학과 교수 (제1저자)
**정희원, 동양대학교 건설공학과 박사과정 (교신저자)
***정희원, 동양대학교 건설공학과 박사과정 (참여저자)
****정희원, 동양대학교 건설공학과 교수 (참여저자)
*****정희원, 한국건설기술연구원 선임연구위원 (참여저자)
*****정희원, 한국건설기술연구원 박사후연구위원 (참여저자)
접수일: 2022년 12월 28일, 수정완료일: 2023년 1월 5일
게재확정일: 2023년 1월 9일

Received: December 28, 2022 / Revised: January 5, 2023
Accepted: January 9, 2023
**Corresponding Author: woals255@dyu.ac.kr
Dept. of Construction Engineering, Dongyang University,
Korea

와 고층건물의 증가에 따른 기존 지하철 구조물에 인접하여 도심도의 굴착 시공이 증가하고 있다[1-5]. 도심지에서의 지반굴착은 흙막이 구조물뿐만 아니라 대규모 수송을 담당하는 지하철 구조물의 안전성 및 열차주행성에 대한 확보가 매우 중요하다[1-5]. 국내 유지관리를 위한 자동화계측 시스템이 지하철 구조물, 터널(고속도로, 고속철도) 등 확대되는 추세이지만 합리적인 실시간 계측을 위한 분석기법 또는 관리기준이 정립되어 있지 않은 실정이다. 현행 기술은 내공변위계, 도상침하계 및 수압계 등을 통해 구조물의 안전성을 평가하고 손상 원인을 해소하는 용도로 사용 중에 있다[1-5]. 현재 지하구조물 및 궤도는 외부요인에 의해 지하구조물의 손상이 다수 발생하는 실정이며 절점에서의 계측결과로 원인을 분석하여 예방차원이 아닌 사후처리에 대해서 측정하고 있는 실정이다[1-5]. 도시의 인구증가 및 고밀화에 따라 기존 구조물에 인접하여 도심도의 굴착 공사 물량이 증가하는 추세인 것으로 분석되었다.[1-5]

물 안전성 예측을 위한 AI기반의 알고리즘 관련 연구가 필요한 실정이다. 종래의 기술은 실시간으로 문제가 발생한 구간을 파악하거나 조치할 수 있는 시간적 여유가 없는 시스템으로 구성되었다. 시간당 하나의 데이터로 24시간 측정하여 1일 최대값을 관리기준치에 비교하여 평가하는 기법이다.

본 연구에서는 실시간으로 측정되는 데이터를 머신러닝 기법을 적용하여 분석을 수행하고 충분한 학습데이터를 바탕으로 측정된 데이터가 안전성을 갖고 향후 데이터가 변화되는 것을 예측할 수 있는 프로그램을 연구하여 종래의 기술보다 확률적으로 높은 안전성을 평가하는 것이 목적이다.

II. 머신러닝 기반의 구조물 안전성

평가 알고리즘 분석

1. 개요

본 연구에서 머신러닝 기법을 적용한 목적은 종래의 기술에 대한 막대한 비용과 시간 및 노력에도 불구하고 기술의 제한적인 한계가 있다. 또한 측정 데이터로 빅데이터를 구축하지 못하였으며, 구조물의 거동을 예측할 수 없는 기법이다.

본 연구에서는 머신러닝을 이용하여 다양한 학습 데이터를 즉, 과거의 데이터들을 적용하여 현재의 데이터에 대한 적정성을 판단하고 향후에 발생가능한 구조물의 거동을 예측하고자 머신러닝 기법을 적용하였다.

본 연구에서 검토한 머신러닝 알고리즘은 선형회귀(Linear regression), 랜덤포레스트(Random forest), 서포트벡터머신(Support vector machine), 다항회귀(Polynomial regression)이다. 본 연구에서는 동일한 학습데이터를 총 4가지의 머신러닝 알고리즘에 적용하여 각각의 알고리즘별 예측결과와 정확성을 실측데이터로 검증하고, 향후 구조물의 관리기준치에 도달하는 시점을 예측하기 위한 분석모델을 학습시켜 그 결과를 분석하였다.

선형회귀 알고리즘은 독립변수와 종속변수의 관계를 설명하는 회귀분석 기법으로 독립변수(X)와 종속변수(Y) 간의 상관관계를 찾기 위한 것이다[6-8]. 회귀분석에 적용되는 독립변수에 따라 단순 선형회귀, 다중 선형회귀로 구분된다[6-8]. 머신러닝에서의 랜덤 포레스트는 분



(a) Photographs of tunnel convergence meter sensor installation (b) Photographs of STL sensor installation



(c) Automated Measurement flow chart



(d) Measured displacement

그림 1. 종래의 자동화 계측 시스템
Figure 1. Conventional automated instrumentation system

시공 및 운영중 발생가능한 다양한 제반조건의 변화에 실시간으로 능동적인 대처가 가능하려면 현장조건 및 다양한 시공정보(재료 및 구조적 특징 포함)에 대한 구조

류 및 회귀 분석 등에 적용되는 앙상블 학습 방법의 구분되며, 학습 과정에서 구성된 다수의 결정 트리로부터 분류 및 평균 예측치(회귀 분석)를 도출한다[9]. 랜덤 포레스트의 큰 특징은 랜덤성(Randomness)에 의하여 각각의 트리들이 다른 특성을 갖는다[9]. 이러한 특성들은 각 트리의 예측(Prediction)을 비상관화 시키며, 결과적으로 성능을 향상시킨다[9]. 또한 랜덤화(Randomization)는 노이즈가 포함된 학습데이터에 대해서도 포레스트가 좋은 성능을 갖는 것으로 분석되었다[9]. 랜덤화는 각 트리들의 학습 과정에서 수행되며, 랜덤 학습데이터 추출 기법을 이용한 앙상블 학습법인 랜덤 노드 최적화(Randomized node optimization) 및 배깅(Bagging)이 주로 사용된다. 이 두 가지 방법은 각각 동시에 적용되어 랜덤화 특성을 크게 향상시킬 수 있다.

서포트 벡터 머신 알고리즘은 패턴 인식 및 자료 분석을 위한 머신러닝 모델로서 Vapnik(1995)이 제시하였다[6,10,11]. 주어진 학습데이터에 대해 규칙 및 데이터의 집합을 찾아내는 알고리즘으로서 주로 분류 및 회귀분석을 위해 사용된다[6,10,11]. 훈련데이터의 특성이 정의하는 변수들로 구성될 때, 학습데이터는 차원 데이터 공간에 분포하고, n차원의 데이터 공간에서 학습데이터 그룹들을 구분하여 최적 경계를 찾아내는 알고리즘이다[6,10,11]. 서포트 벡터 머신 알고리즘의 목적은 n차원의 데이터 공간에서 학습데이터의 그룹들을 분류해 내는 최적의 분할선(Optimal decision boundary)을 분석하는 것이다. 데이터 공간에서 학습데이터 그룹을 분류할 수 있는 초평면(Hyperplane)은 수없이 많으며, 초평면에 의하여 분리된 학습데이터들 중 초평면과 가장 가까운 그룹과 초평면 사이의 거리를 마진(Margin)이라고 정의한다.

2. 머신러닝 기법을 적용한 구조물 안정성 분석

자동화계측을 통해 획득한 시계열 데이터는 그림 2와 같다. 자동화계측을 통해 일정 기간동안 일정한 시간간격으로 수집되는 계측데이터를 시간의 경과에 따라 정리하여 시계열 분석을 수행할 수 있다[1-5]. 시계열 데이터의 예측은 불확실성이 내포되어있으며 데이터에 대한 신뢰성을 확보해야 한다. 여러 패턴이 복잡하게 섞여 있으며 전체적인 데이터가 일정하게 증가 또는 감소하는 패턴을 나타내는 것으로 분석되었다. 또한 데이터의 표본이 충분하지 않거나 데이터 노이즈 등 다양한 변수가 발생 가능한 데이터이다.

각 알고리즘을 통해 구조물 관리기준치를 달성하기까지의 예측 시점을 분석하였으며, 실제 구조물에서 측정된 구조물의 변위 데이터를 누적시켜 머신러닝의 학습데이터 세트에 적용한 시계열 데이터는 그림 2와 같다.

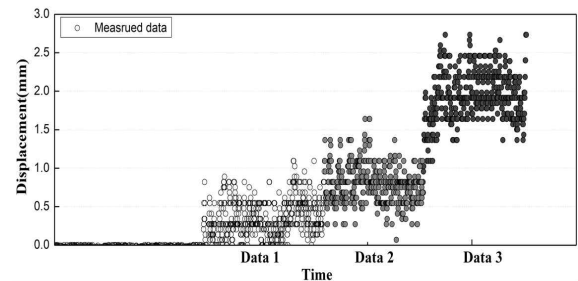


그림 2. 자동화계측결과(시계열 측정데이터)
Figure 2. Automated measured data(Time-history data)

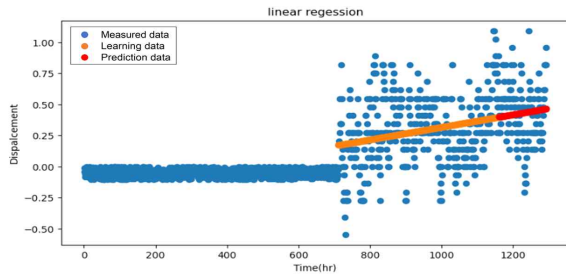
본 연구에서는 그림 2와 같이 공용중인 도시철도 지하철 터널구조물에서 3년간 한시간에 하나의 데이터를 측정하여 2,261개의 데이터셋을 확보하였으며, 머신러닝 알고리즘의 학습데이터로 적용하였다.

훈련데이터는 그림 2와 같이 변위차가 발생하는 단계별로 크게 3가지 구간(Data 1~3)으로 구분하였다. Data 1은 0hour~1,293hour의 측정데이터이며, Data 2는 0hour~1,764hour의 측정데이터, Data 3는 0hour~2,261hour의 측정데이터로 구분하여 학습데이터로 적용하였다. 본 연구에서는 자동화계측의 대상구조물인 터널 구조물의 실시간 거동특성이 외적 요인(인접굴착 등)으로 인해 변화되는 것을 파악하고자 터널의 내공변위를 측정하여 훈련데이터로 적용하고 머신러닝 알고리즘을 통해 구조물의 안전성 평가기준에 도달하는 시점을 예측하였다.

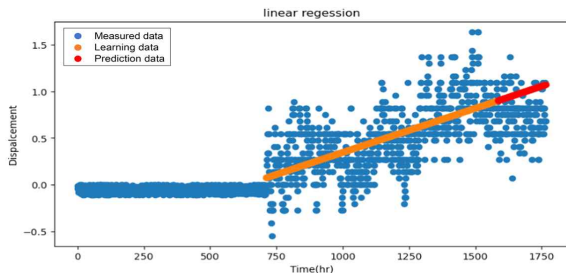
본 연구에서는 다변량 시계열 데이터를 분석과 예측을 통해 문제를 다루고 순차적으로 관찰된 수 많은 변수들에 대해 학습하고자 한다. 또한 Data별 측정데이터의 90%까지 해당하는 데이터를 머신러닝 알고리즘 기법에 적용하였다. 학습한 결과를 통해 터널 관리기준치에 도달하는 시점을 예측하는 결과는 그림 3 ~ 그림 5에 나타내었다.

선형회귀 분석은 이미 알고 있는 X, Y값을 통해 머신러닝의 Input으로 적용하여 훈련시킴으로서 손실을 최소화하고 식(1)에서 β 와 b 의 두 개의 Parameters의 값을 구하는 것이 선형회귀의 목표이다.

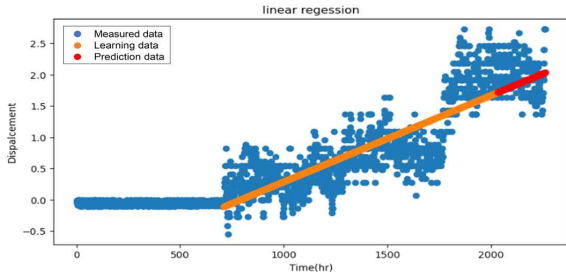
구조물의 변위 거동이 발생하는 722시간대부터 2,035시간의 데이터를 학습시켜 구조물의 관리기준치에 도달하는 시점을 예측하였다.



(a) Data 1 (0hr~1293hr)



(b) Data 2 (0hr~1764hr)

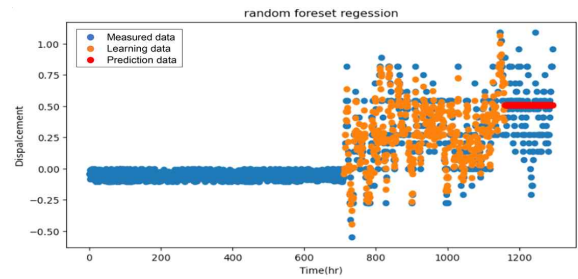


(c) Data 3 (0hr~2261hr)

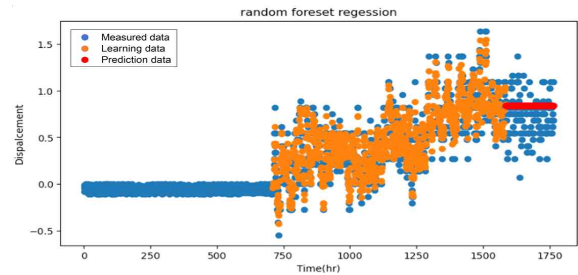
그림 3. 선형회귀 알고리즘 분석결과
Figure 3. Result of linear regression

분석결과, 그림 3과 같이 선형회귀 알고리즘의 적정성은 Data 1의 예측치 결과(0.5mm)와 Data 2의 확보된 데이터를 통한 학습결과의 시작점(0.6mm)과 비교 시 약 83%의 유사성을 나타낸 것으로 분석되었으며, Data 2의 예측치 결과(1.0mm)와 Data 3에서 확보된 데이터를 통한 학습결과의 시작점(1.3mm)과 비교 시 약 76%의 예측 데이터의 정확성이 분석되었다.

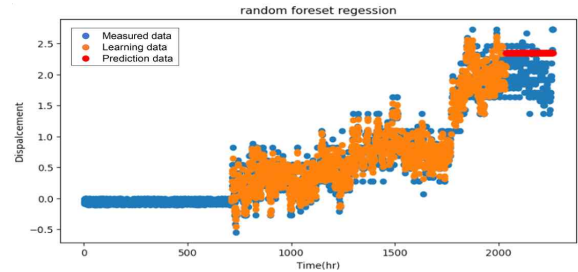
선형회귀 알고리즘은 학습구간에 맞추어 수행하므로 관리기준치에 도달하는 시점의 예측 결과가 선형적으로 증가하는 경향으로 분석되었다.



(a) Data 1 (0hr~1293hr)



(b) Data 2 (0hr~1764hr)

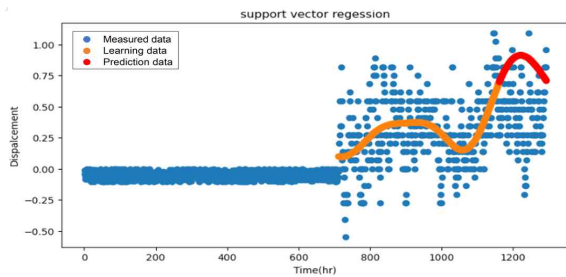


(c) Data 3 (0hr~2261hr)

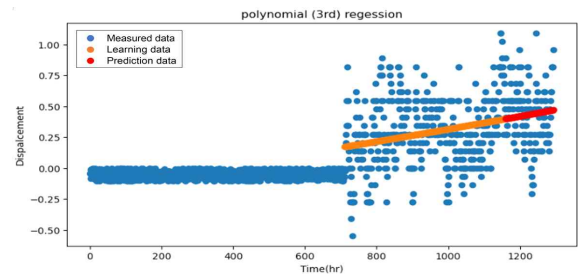
그림 4. 랜덤포레스트 알고리즘 분석결과
Figure 4. Result of random forest

분석결과, 그림 4와 같이 랜덤포레스트 알고리즘은 Data 1의 예측치 결과(0.5mm)와 Data 2에서 학습결과의 시작점(0.7mm)과 비교 시 약 70%의 데이터의 유사성을 나타낸 것으로 분석되었으며, Data 2의 예측치 결과(0.7mm)와 Data 3에서 학습결과의 시작점(1.0mm)과 비교 시 약 70%의 예측데이터의 정확성이 분석되었다.

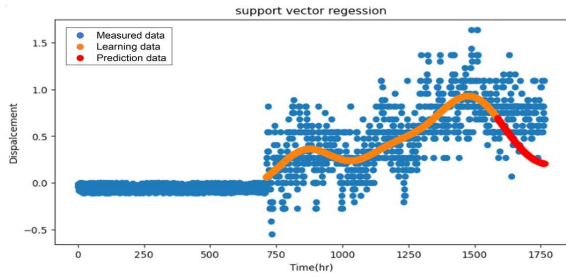
랜덤 포레스트는 기존 학습데이터에 대한 평균치를 나타내고, 학습데이터 이후 구조물의 거동을 반영을 하지 못하고 과대, 과소평가를 할 수 있는 여지가 있는 것으로 분석되었다.



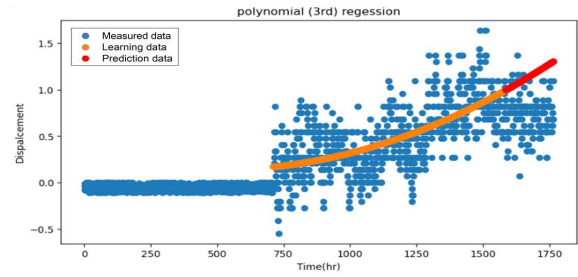
(a) Data 1 (0hr~1293hr)



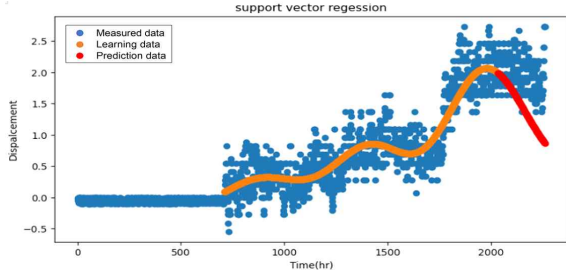
(a) Data 1 (0hr~1293hr)



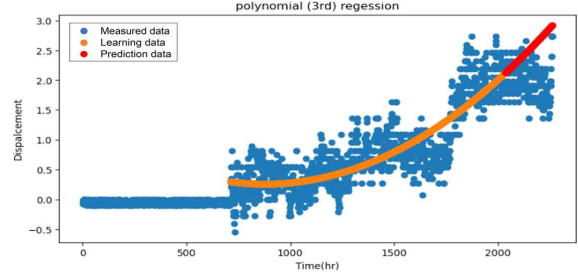
(b) Data 2 (0hr~1764hr)



(b) Data 2 (0hr~1764hr)



(c) Data 3 (0hr~2261hr)



(c) Data 3 (0hr~2261hr)

그림 5. 서포트벡터머신(SVM) 알고리즘 분석결과
 Figure 5. Result of support vector machine

그림 6. 다항회귀 알고리즘 분석결과
 Figure 6. Result of polynomial regression

서포트벡터머신 알고리즘 분석결과, 그림 5와 같이 n 개의 변수들을 n 차원 데이터 공간 안에 변수들을 묶어 구분해 내는 최적의 분할선을 찾아내는 것이나 Data 1의 예측치 결과(0.75mm)와 Data 2에서 학습결과의 시작점(0.6mm)과 비교 시 약 80%의 데이터의 유사성을 나타낸 것으로 분석되었으며, Data 2의 예측치 결과(0.25mm)와 Data 3에서 학습결과의 시작점(1.5mm)과 비교 시 약 17%의 예측데이터의 정확성이 분석되었다. 또한 관리기준치에 도달하는 예측에 있어 정확성이 높지 않으며, 경향이 반대로 나오는 것으로 분석되었다.

다항 회귀 알고리즘을 적용할 때 주의할 점은 과대적합(Overfitting)과 과소적합(Underfitting)이다. 과대적합은 치수가 너무 높을 경우 학습 데이터의 오차는 감소하나 실제 데이터가 Input으로 적용됐을 때 오차가 증가하는 문제가 발생하고, 과소적합은 치수가 낮을 경우 학습이 제대로 이뤄지지 않아 적합성이 떨어진다.

다항회귀 분석결과, 그림 6과 같이 Data 1의 예측치 결과(0.5mm)와 Data 2에서 학습결과의 시작점(0.6mm)과 비교 시 약 80%의 데이터의 유사성을 나타낸 것으로 분석되었으며, Data 2의 예측치 결과(1.25mm)와 Data 3에서 학습결과의 시작점(1.5mm)과 비교 시 약 83%의 예측데이터의 정확성이 분석되었다.

Data별 머신러닝 알고리즘 분석결과, 학습데이터셋이 증가할수록 다항회귀 알고리즘이 제일 정확성이 높고 오차의 범위가 줄어드는 것으로 분석되었다. 또한 구조물의 변위가 발생하는 구간에서의 데이터를 분석하여 관리기준치에 도달하는 예측이 잘 이뤄진 것으로 분석되었다.

III. 결론

본 연구에서는 종래의 계측결과보다 실시간으로 측정 데이터를 분석하고 관리기준치에 도달하는 예측 시점을

확인할 수 있는 분석 알고리즘을 개발하고자 실제 구조물의 변위 데이터를 머신러닝을 적용하여 분석하였다.

(1) 랜덤포레스트 알고리즘과 서포트벡터머신 알고리즘 적용 결과, 학습데이터 이후의 구조물의 거동을 반영을 하지 못하고 과소, 과대평가를 할 수 있는 소지가 있는 것으로 분석되었다.

(2) 머신러닝 모델은 학습된 데이터를 기반으로 예측을 수행하기 때문에 다양한 변수 및 조건에서 획득한 데이터가 다양하게 존재할수록 모델의 예측 성능이 향상된다. 반면, 데이터의 표본이 적은 경우는 예측 성능이 감소하거나, 학습결과가 과적합 또는 편향성이 증가하게 되는 것으로 분석되었다.

(3) 선형회귀 알고리즘과 다항회귀 알고리즘에서 분석한 데이터셋에서 구조물 관리기준치에 도달하는 예측에 적합한 모델은 다항회귀 알고리즘인 것으로 분석되었다. 추후 보다 다양한 구조물을 대상으로한 시계열분석이 가능한 자동화측측 데이터를 확보하여 본 연구에서 도출한 머신러닝 알고리즘의 신뢰성 향상을 위한 추가적인 연구가 필요하다.

(4) 본 연구에서 제시한 구조물 안전성 분석을 위한 머신러닝 알고리즘을 적용함으로써 실시간으로 측정된 구조물의 응답치를 학습데이터로 적용하여 거동특성을 분석하고, 이를 바탕으로 임의의 시점에서의 구조물의 거동 또는 구조적 안전성을 예측하는 데에 활용할 수 있을 것으로 판단된다.

References

[1] J.Y. Choi, H.H. Lee, Y.S. Kang, J.S. Chung, "Evaluation of Structural Stability of Tunnel due to Adjacent Excavation on Urban Transit", Journal of the Convergence on Culture Technology(JCCT), Vol. 6, No. 2, pp. 503-508, 2020.

[2] J.Y. Choi, S.H. Kim, H.H. Lee, J.S. Chung, "Improvement of Automatic Measurement Evaluation System for Subway Structures by Adjacent Excavation", Materials, Vol. 14, No. 24, pp. 1-20, 2022.

[3] J.Y. Choi, G.N. Yang, T.J. Kim and J.S. Chung, "Analysis of Ground Subsidence according to Tunnel Passage in Geological Vulnerable Zone", Journal of the Convergence on Culture Technology(JCCT), Vol. 6, No. 3, pp. 393-399,

2020.

<http://doi.org/10.17703/JCCT.2020.6.3.393>

[4] J.Y. Choi, G.N. Yang, T.J. Kim and J.S. Chung, "Analysis of Changes in Groundwater Level according to Tunnel Passage in Geological Vulnerable Zone", Journal of the Convergence on Culture Technology(JCCT), Vol. 6, No. 3, pp. 369-375, 2020.
<http://doi.org/10.17703/JCCT.2020.6.3.369>

[5] J.Y. Choi, S.I. Cho and J.S. Chung, "Parameter Study of Track Deformation Analysis by Adjacent Excavation Work on Urban Transit", Journal of the Convergence on Culture Technology(JCCT), Vol. 6, No. 4, pp. 669-675, 2020.

[6] T.H. Kang, S.W. Choi, C.H. Lee, S.H. Jang, "A Study on Prediction of EPB shield TBM Advance Rate using Machine Learning Technique and TBM Construction Information", Journal of Korean Tunnelling and Underground Space Association, Vol. 6, No. 6, pp.540-550, 2020.

[7] David A. Freedman, Statistical Models: Theory and Practice, Cambridge University Press, pp. 458, 2009.

[8] Rencher, Alvin C, Christensen William F, Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, pp. 800, 2012.

[9] Breiman L, Random Forests. Machine Learning, Kluwer Academic Publishers, pp. 5-32, 2001.

[10]Cortes, C.; Vapnik, V. Machine Learning, Support-vector networks, Kluwer Academic Publishers, pp. 273-297, 1995.

[11]Vapnik, V, The nature of statistical learning theory, Springer, pp. 314, 2000.

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2022-0-00829, 스킨 센서와 A.I.를 활용한 SOC 시설물 실시간 이상 감지 시스템 개발)