

WiFi 신호강도 기반의 실내 측위를 위한 머신러닝 회귀 알고리즘의 비교 분석

Comparative Analysis of Machine Learning Regression Algorithms for Wi-Fi Signal Strength-Based Indoor Positioning

김용현¹⁾

Kim, Yonghyun

Abstract

Wi-Fi strength-based indoor positioning remains widely used across various fields because it does not require additional signal transmitters or receivers. A pre-constructed strength database of radio signals is necessary for Wi-Fi signal strength-based positioning, typically structured as a tabular dataset. Despite the rapid advancements in artificial intelligence, applying deep learning to such tabular data still results in unstable learning, with traditional machine learning-based regression algorithms outperforming in terms of positioning accuracy. In this paper, we conduct an in-depth analysis of these trends by applying various machine learning regression algorithms to a carefully constructed signal strength distribution dataset, comparing their strengths and weaknesses. The algorithms used include Extra Trees Regression, Random Forest, Support Vector Regression, and Extreme Gradient Boosting. Each algorithm's accuracy and training time were analyzed. The experimental results show that Extra Trees Regression achieved the highest positioning accuracy, while XGBoost demonstrated the fastest training time. The dataset used in this study integrates data from all floors of a single building, organized into x-, y-, and z-(floor) coordinates. Most machine learning-based regression algorithms have demonstrated the ability to predict the x- and y-coordinates with a root mean square error of less than 2 meters, while achieving over 99% accuracy in predicting the z-coordinate(floor level).

Keywords : WiFi-based positioning, Indoor positioning, Machine learning, Regression, Tabular dataset, Deep learning, Ensemble learning

초 록

WiFi 신호강도 기반 실내 위치 결정은 추가적인 전파 신호의 송수신기를 설치하지 않아도 사용할 수 있다는 장점 덕분에 여전히 다양한 분야에서 활용되고 있다. WiFi 신호강도 기반 측위를 위해서는 사전에 WiFi 신호의 전파강도 분포도가 데이터베이스로 구축되어 있어야 하며, 이러한 데이터는 실제 테이블 형태로 저장된다. 최근 인공지능 기술이 급격히 발전했음에도 불구하고, 이러한 테이블 데이터에 딥러닝을 적용하면 학습이 불안정하고, 측위 정확도 성능은 여전히 기계학습 기반의 회귀 알고리즘이 우수한 것으로 알려져 있다. 본 논문에서는 정교하게 구축된 WiFi 신호강도 분포도를 기반으로 다양한 기계학습 기반 회귀 알고리즘을 적용하여 각 방법론의 장단점을 비교 분석하였다. Extra Tree Regression, Random Forest, Support Vector Regression, Extreme Gradient Boosting 알고리즘을 적용하였으며, 각 알고리즘의 정확도 성능과 학습 시간을 분석하였다. 실험 결과, 측위 정확도 성능은 Extra Tree Regression이 가장 우수하였으며, 학습 속도는 Extreme Gradient Boosting이 가장 빠른 것으로 나타났다. 본 연구에서 사용된 데이터는 하나의 건물 내 모든 층을 통합하여 x축 좌표, y축 좌표, z축(층)으로 구성되었으며, 대부분의 기계학습 기반 회귀 알고리즘이 평균제곱근 오차 2m 이내로 x축 좌표, y축 좌표를 예측하였을 뿐만 아니라 z축(층)은 99% 이상의 정확도로 예측하는 것을 확인하였다.

핵심어 : WiFi 신호 기반 위치 결정, 실내 위치 결정, 기계학습, 회귀, 테이블 데이터 세트, 딥러닝, 앙상블 학습

Received 2024.09.25, Revised 2024.10.22, Accepted 2024.10.30

1) Corresponding Author, Member, Mobility UX Research Section, Electronics and Telecommunications Research Institute (E-mail: yhkeen@etri.re.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

GNSS (Global Navigation Satellite System)를 활용할 수 없는 실내 환경에서 정확한 위치 추정은 재해 재난 대피와 같은 긴급한 사항뿐만 아니라 물류센터의 재고관리, 실내 내비게이션, 위치기반 마케팅 및 제품 상용화 활용 등 다양한 분야에서 매우 중요한 과제로 여겨지고 있다(Gu *et al.*, 2009; He and Chan, 2015; Kushki *et al.*, 2007). 특히 실내의 공간의 연속적 위치 결정 측면에서 GNSS 위성 신호의 수신에 제한되는 실내 공간에서 사용자 또는 로봇의 위치를 결정하는 수요는 지속적으로 증가하고 있는 상황에 있다. 이와 같은 수요에 맞춰 WiFi, BLE (Bluetooth Low Energy), UWB (Ultra-WideBand), RFID (Radio Frequency IDentification) 등을 활용하는 다양한 무선 신호기반 측위 알고리즘들이 제안되었지만(Jiang *et al.*, 2021; Nosrati *et al.*, 2022), 전파 신호의 복잡한 다중 경로 효과, 신호 간섭에 영향을 받는 특성에 따라 각각의 장단점을 갖고 활용되고 있다(Alarifi *et al.*, 2016; Bianchi *et al.*, 2018).

다양한 실내 위치 추정 기술 중 WiFi 신호를 이용한 실내 위치 결정은 건물 구조, 장애물 등의 복잡성으로 인해 신호가 불안정할 수 있다는 단점이 있지만, 별도의 인프라 하드웨어 설치 없이 기존에 구축된 네트워크를 사용할 수 있다는 점에서 널리 사용되고 있다(Liu *et al.*, 2007). 즉, UWB 및 의사위성(Pseudolite)을 이용한 실내 위치 추정을 위해서는 추가로 전파 송수신 앵커(anchor)를 설치하여야 하는 데 비해, 대부분의 실내 환경에 이미 WiFi 네트워크가 구축되어 손쉽게 이용할 수 있는 점이 가장 큰 장점이다(Zafari *et al.*, 2019). 이러한 이유로, WiFi 기반의 실내 측위 기술은 비용 효율성이 높으며, 최근 기계학습 및 딥러닝 기반의 다양한 방법론과 결합해 활용되고 있다(He and Chan, 2015).

WiFi 신호 기반의 실내 측위를 위해서는 사전에 신호강도 라디오(radio) 맵이 구축되어 있어야 하며, 이와 같은 데이터는 실제 테이블 형태의 데이터로 인식된다. 하지만, 이러한 테이블형 데이터에 딥러닝 알고리즘을 적용하는 경우, 학습이 불안정해지는 문제가 발생할 수 있다(Grinsztajn *et al.*, 2022; Shwartz-Ziv and Armon, 2022). 특히, 테이블 형태의 데이터에 적용한 딥러닝 기반 알고리즘의 성능은 때때로 기계학습 기반의 회귀 알고리즘보다 저조한 것으로 알려져 있으며, 실내 측위의 정확도 면에서도 기계학습 기반의 회귀 알고리즘이 더 우수하다는 연구 결과들이 보고되고 있다(Ji *et al.*, 2021; Nessa *et al.*, 2020). 이는 주로 WiFi 신호의 복잡한 변동성을 모델링하는 데 있어 딥러닝 알고리즘이 가진 한계로 해석

될 수 있는 측면이 있지만, 다량의 학습데이터를 요구하는 딥러닝 관점에서 학습데이터가 부족하다는 의미일 수도 있다.

본 연구에서는 WiFi 신호를 이용한 신호강도 기반의 실내 측위에서 다양한 기계학습 기반 회귀 알고리즘을 적용하여 그 성능을 정량적으로 비교·분석하였다. 정교하게 구축된 WiFi 신호강도 데이터 세트를 바탕으로 각 회귀 알고리즘의 학습 및 추론 시간을 분석하였으며, 이를 통해 실제 실내 측위 시스템으로의 적용 가능성을 평가하였다. 특히, 각 알고리즘이 x - 및 y -좌표와 z (층) 좌표를 얼마나 정확하게 예측하는지에 중점을 두고 분석하였으며, 궤적 기반의 학습데이터를 이용하여 각 궤적이 어떻게 변하는지 시각화하여 해석하였다.

본 논문의 구성은 다음과 같다. 1장에서 연구 배경과 기존 연구의 한계점을 기술하였으며, 2장에서 WiFi 기반 실내 측위 데이터 세트와 측위 원리를 기술하였다. 3장에서는 데이터 전처리 과정과 함께 본 연구에서 사용된 기계학습 기반 회귀 알고리즘의 적용 방법을 기술하였다. 4장에서는 실험 결과와 그에 대한 분석을 제공하며, 다양한 회귀 알고리즘의 성능을 정량적으로 평가하고, 시각화를 통해 예측된 궤적과 실제 데이터를 비교하였다. 마지막으로 5장에서는 본 연구의 의의를 요약하고 결론을 도출하였다.

2. WiFi 기반 실내 측위 데이터

2.1 WiFi 신호강도 기반 실내 측위

일반적으로 실내에서의 WiFi 무선 신호 수신 강도는 AP (Access Points)와 휴대전화 단말기의 거리 증가에 따른 신호강도 감쇠, 그리고 벽 및 기타 실내 구조물을 통과할 때 발생하는 신호 감쇠와 음영 감쇠(shadow fading) 효과를 포함하는 경로 손실 모델로 설명될 수 있다(Shang *et al.*, 2014). 음영 감쇠 효과는 무선 통신에서 발생하는 신호의 감쇠 현상 중 하나로, 주로 건물, 벽, 실내 가구와 같은 큰 장애물에 의해 신호가 차단되거나 감쇠되는 현상을 의미한다. 특정 공간의 좌표 (x, y, z) 에서 AP로부터 수신된 신호강도 RSSI (Received Signal Strength Indicator)는 다음의 Eq.(1)과 같이 나타낼 수 있다.

$$RSSI(x, y, z)[dBm] = -10n \log_{10} d + \alpha \quad (1)$$

여기서, n 은 경로감쇠지수, d 는 AP와 단말기 사이의 거리, α 는 자유공간의 참조거리에서 측정된 단말기 수신신호강도를 의미한다. 통상적으로 완전한 신호강도일 때, 약 -30 dBm을 가지며, -80 dBm 이하일 경우 통신연결이 불안정하거나 통

신두절이 발생한다. WiFi를 이용한 RSSI 신호강도 데이터 세트는 특정한 위치에서 측정된 AP들의 신호강도 세기로 구성되어 있으며, 일반적으로 위치 참조점(reference point) 좌표 (x, y, z) 와 신호강도 값(dBm)이 행으로 구성된 테이블 형태로 저장된다.

일반적으로 WiFi 기반의 실내 위치 결정은 Fig. 1과 같이 온라인 단계와 오프라인 단계로 구분할 수 있다. 먼저, 오프라인 단계에서는 실내 측위를 위해 참조점에서 WiFi 신호강도를 측정해 데이터베이스화하는 단계를 수행하며, 통상적으로 촘촘한 격자 형태의 데이터를 구축한다. 또한, 오프라인 단계에서는 기계학습 또는 딥러닝 기반의 다양한 알고리즘의 학습을 수행하여 모델을 개발한다. 마지막으로, 온라인 단계에서는 오프라인 단계에서 개발된 모델이 서버에 배포되어, 측위 단말기 사용자, 로봇 등의 측위 요청에 따른 위치 좌표를 전송하는 단계로 구성되어 있다.

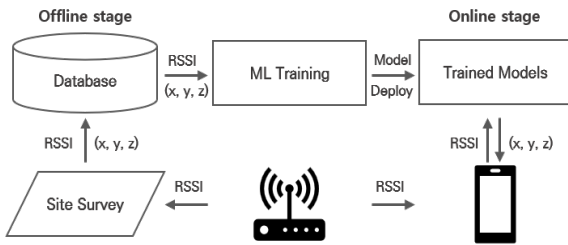
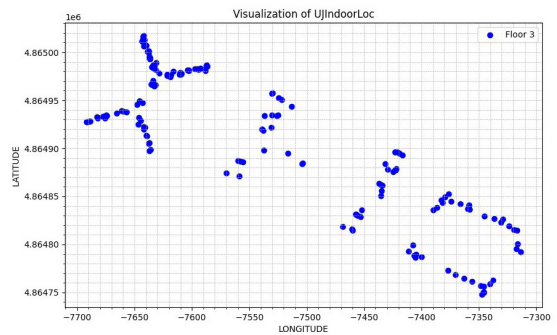


Fig. 1. Systematic flow of WiFi-based indoor positioning

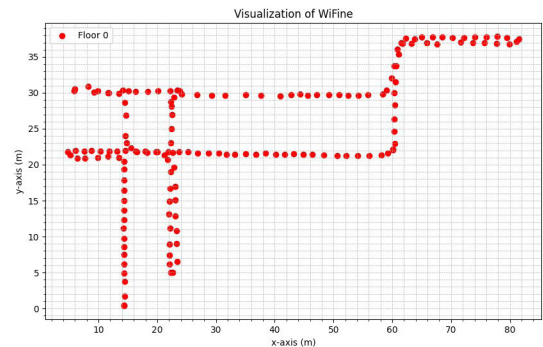
2.2 공개 데이터 세트 분석

WiFi 기반의 실내 측위를 위해서는 사전에 참조점과 WiFi RSSI를 기반으로 하는 데이터 세트의 구축이 선행된다. 최근까지 다양한 측위 알고리즘의 적용 및 분석을 위해 다양한 데이터 세트가 공개되었으며, 이와 같은 데이터 공유는 연구 결과의 재생산을 위해서 매우 중요한 과정이라 할 수 있다. 본 연구에서는 2014년 이후에 공개된 주요 데이터 세트를 조사하였으며, 데이터 세트로 구성된 건물의 개수($N_{Building}$), 전체 층(N_{Floor}), 참조점(N_{RP}), AP 수(N_{AP}), 궤적 수(N_{Traj})를 Table 1에 정리하였다. WiFine라는 명칭으로 공개된 데이터 세트는 2만 6천개가 넘는 참조점으로 타 데이터 세트에 비해 월등히 많은 데이터를 보유하고 있으며, 대부분의 데이터 세트가 시계열 궤적 정보를 포함하고 있지 않은 데 반해, 290개의 궤적 정보를 포함하고 있음을 알 수 있다(Khassanov et al., 2021). 그 외에 실내 위치 측위와 관련된 주요 학회인 IPIN (International Conference on Indoor Positioning and Indoor Navigation)에서 2016년 이후 공개된 데이터 세트에 일부 궤

적 정보가 포함되어 있지만, 딥러닝 관점에서의 비선형 함수 근사, 데이터 분포의 일반화 측면에서 시계열 예측을 적용하기에 데이터가 부족함을 확인할 수 있다(Liu et al., 2024). 즉, 휴대전화 등의 측위 단말기를 소지한 사람 또는 로봇의 이동에 따른 궤적 정보는 본질적으로 시계열 데이터로 각 시점의 위치와 RSSI는 직전 위치와 RSSI가 영향을 미친다는 시간적 의존성을 반영해야 한다. 최근 LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units) 등의 순환 신경망 구조를 활용한 측위 연구가 제안되었으며(Liu et al., 2024), 이와 같은 신경망 구조는 잠재 공간(latent space)에서 시·공간적 데이터의 의존성을 효과적으로 학습할 수 있는 것으로 알려져 있다. Fig. 2에 각각 UJIndoorLoC 데이터와 WiFine 데이터라는 이름으로 공개된 데이터의 일부를 나타내었다. Fig. 2에서 확인할 수 있듯이, UJIndoorLoC 데이터는 시계열 궤적 정보를 포함하지 않고 정적으로 단순히 건물 층의 전체 공간을 포함하려는 측면에서 구축되었다면, WiFine 데이터는 건물 복도 공간을 사람이 이동하며 데이터가 구축된 측면이 주요한 차이점이라 할 수 있다.



(a) UJIndoorLoC Data
(European Petroleum Survey Group:3857)



(b) WiFine Data

Fig. 2. Indoor positioning dataset

Table 1. Opened WiFi positioning dataset

Dataset	$N_{Building}$	N_{Floor}	N_{RP}	N_{AP}	N_{Traj}
UJIndoorLoc	3	2	933	520	N/A
XJTUIndoorLoc	1	2	969	515	N/A
UTSIndoorLoc	1	16	1,840	589	N/A
Tampere	1	5	4,648	991	N/A
Library	1	2	N/A	448	N/A
JUIndoorLoc	1	3	1,000	172	N/A
IPIN2016	4	6	2,007	N/A	26
IPIN2017	3	6	2,697	N/A	38
IPIN2018	1	3	N/A	N/A	38
WiFine	1	3	26,418	436	290

3. 머신러닝 회귀를 이용한 실내 측위

3.1 사용 데이터 및 전처리

본 연구에서는 정적 측위, 시계열 예측 모두 적용할 수 있는 WiFine 데이터 세트를 이용하여 기계학습 기반의 여러 회귀 알고리즘을 적용하였다. 대부분의 연구에서 간헐적으로 기록된 AP들이 특정한 개수 이하일 경우 해당 AP를 제외하는 임계치를 적용해 제외하는 데 반해, 본 연구에서는 각 알고리즘의 강건성(robustness)을 평가하기 위해 임계치 적용을 생략하였다. 즉, 실제 WiFi 측위 환경에서는 신호가 강한 AP만이 연결되는 것이 아니라 신호가 약하며 간헐적으로 연결되는 AP들이 존재하기 때문에 해당 환경을 가정해 강건성을 평가하였다. WiFine 데이터 세트는 x, y, z축 좌표와 3개 층에 대한 클래스로 구성되어 있으며, 본 연구에서는 x, y축 좌표와 3개 층에 대한 클래스를 예측하는 것을 목적으로 해당 데이터 세트를 활용하였다. 마지막으로, 음수로 되어있는 RSSI dBm 값을 0~1 사이의 값으로 정규화 시키기 위해 다음의 정규화 식을 적용하였다(Liu *et al.*, 2024).

$$RSSI_{Normalized} = \frac{RSSI - RSSI_{min}}{RSSI_{max} - RSSI_{min}} \quad (2)$$

여기서, $RSSI$ 는 각 AP의 RSSI 값을 지칭하며, WiFine 데이터 세트의 경우 -100부터의 음수로 기록되어 있다. 따라서, $RSSI_{min}$ 은 -100으로 적용하였고 $RSSI_{max}$ 는 각 AP의 최대 신호강도로 정의된다.

3.2 기계학습 기반 회귀 알고리즘

본 연구에서는 RFR (Random Forest Regression), SVR

(Support Vector Regression), XGBoost (Extreme Gradient Boosting), ETR (Extra Trees Regressor) 알고리즘을 실내 위치 결정 문제에 적용하였다. 각 모델은 파이썬 기반의 Scikit-learn 라이브러리를 이용하여 구현되었으며, 학습 모델은 효율적인 저장과 로딩을 위해 pickle 형식으로 저장되었다. 학습된 모델은 별도의 테스트 데이터에 대해 추론을 수행하였고, 예측 결과를 바탕으로 성능을 비교하고 분석하였다. 또한, 결과의 해석을 위해 각 모델의 평균 제곱근 오차(Root Mean Square Error, RMSE) 성능 지표를 활용하였으며, 실제값과 예측값 간의 차이를 시각화하여 모델의 예측 성능을 평가하였다.

3.2.1 Random Forest Regression

RFR은 앙상블(ensemble) 학습 방법의 하나로, 여러 개의 의사결정 트리를 구성하여 회귀 예측 결과를 종합함으로써 예측 정확도를 높이고 과적합(overfitting)을 최소화하는 방법론이다(Ji *et al.*, 2021). RFR은 다수의 트리를 독립적으로 학습시키고 이들의 예측값을 결합하여 성능을 향상시킨다. RFR 알고리즘은 훈련 데이터에서 무작위로 표본을 추출하고, 트리 학습 시 일부 특징을 무작위로 선택하여 트리 간 상관성을 줄이고 모델의 일반화 성능을 높이는 전략을 사용한다. RFR은 크게 네 단계로 구성된다. 첫째, Bootstrapping 단계에서는 중복을 허용하여 훈련 데이터에서 무작위 표본을 추출한 후, 이를 바탕으로 여러 트리를 생성한다. 각 트리는 서로 다른 표본을 기반으로 학습되어 다양한 패턴을 학습할 수 있다. 둘째, Random Feature Selection 단계에서는 각 노드에서 일부 특징을 무작위로 선택하여 최적의 분할(subset)을 찾는다. 이를 통해 트리 간의 상관성을 줄이고, 다양한 의사결정 트리를 생성함으로써 모델의 강건성을 높인다. 셋째, 트리 구성 단계에서는 각 트리가 Bootstrapping 표본과 선택된 특징을 바탕으로 독립적으로 학습된다. RFR은 일반적으로 가지치기(pruning)를 수행하지 않아 트리들이 깊고 복잡하게 성장할 수 있도록 허용한다. 마지막으로, Aggregation 단계에서는 각 트리의 예측값을 평균하여 최종 회귀 예측값을 도출한다. 이 과정은 예측의 변동성을 줄이고, 보다 안정적이고 정확한 결과를 도출하는 데 기여한다. RFR은 특히 비선형 문제에서 뛰어난 성능을 보이며, 여러 트리의 평균을 취함으로써 과적합을 방지하고 예측의 안정성을 높인다. 또한, 특징의 중요도를 계산할 수 있어 모델의 해석 가능성이 높으며, 이는 해석력이 부족한 딥러닝 모델에 비해 중요한 장점으로 작용한다. 특히, RFR은 비모수적(non-parametric) 특성으로 인해 데이터의 분포에 대한 가정이 필요하지 않아 다양한 형태의 데이터에 적용할 수 있다.

3.2.2 Support Vector Regression

SVR은 SVM (Support Vector Machine)의 개념을 회귀 문제에 적용한 알고리즘으로, 데이터를 고차원 공간으로 변환하여 최적의 초평면(hyperplane)을 찾고 이를 통해 회귀 예측을 수행하는 방법이다(Jondhale *et al.*, 2022). SVR은 회귀 문제에서 가능한 적은 오차로 데이터를 설명하는 최적의 함수(초평면)를 찾는 것을 목표로 한다. 즉, SVR의 핵심은 주어진 허용 오차(epsilon) 내에서 예측된 값이 포함되는 결정 경계(decision boundary)를 설정하고, 이를 최소화하면서 예측에 필수적인 소수의 support vector만을 선택하여 최적화 문제를 해결하는 방법이다. SVR 알고리즘은 크게 세 단계로 이루어진다. 첫째, 커널 함수(kernel function)를 사용하여 데이터를 고차원 특징 공간으로 변환한다. 이를 통해 선형적으로 구분되지 않는 데이터를 고차원에서 선형적으로 구분할 수 있으며, 자주 사용되는 커널로는 RBF(Radial Basis Function) 커널과 다항식(Polynomial) 커널이 있다. 둘째, 고차원 공간에서 설정한 오차 허용 범위 안에 있는 데이터는 무시하고, 범위를 벗어난 데이터에 대해서만 Penalty를 부여한다. 이 Penalty는 과적합을 방지하고 모델의 일반화 성능을 높이는 데 기여하며, 하이퍼파라미터를 통해 조정된다. 셋째, SVR은 결정 경계를 형성하는 데 필요한 소수의 데이터 포인트(support vector)만을 사용하여 모델을 훈련한다. 이에 따라 데이터의 양이 많더라도 support vector만으로 효율적인 모델 구성이 가능하다. SVR은 비선형 데이터에서도 우수한 성능을 보이며, 특히 커널 트릭을 사용하여 복잡한 데이터 패턴을 효과적으로 학습할 수 있다. 또한, 유연한 오차 허용 범위를 설정할 수 있어 과적합을 방지하고 모델의 일반화 성능을 높이는 데 도움을 준다. 하지만 SVR은 대규모 데이터 세트에서 훈련 시간이 오래 걸릴 수 있으며, 커널 선택과 하이퍼파라미터 설정에 크게 의존해, 잘못된 커널 또는 하이퍼파라미터 설정은 성능 저하로 이어질 수 있으므로 신중한 선택이 필요하다.

3.2.3 XGBoost

XGBoost는 의사결정 트리 기반의 Boosting 기법을 확장한 고성능 알고리즘이다(Singh *et al.*, 2022). XGBoost는 회귀 및 분류 문제에서 뛰어난 성능을 발휘하며, 특히 대규모 데이터 세트에 대한 빠른 학습 속도와 높은 정확성으로 주목받았다. XGBoost 알고리즘은 기존 Gradient Boosting 알고리즘을 개선하여 병렬 처리, 과적합 방지, 메모리 최적화 등의 성능 향상을 이루었으며, 크게 다음의 네 단계로 구성된다. 첫째, 기존 모델의 잔차(Residual)를 기반으로 새로운 의사결정 트리를 추가하여 모델을 개선하는 방식이며, 각 트리는 이전 모델

이 예측하지 못한 오차를 학습하여 성능을 점진적으로 향상시킨다. 둘째, 가중치 업데이트 단계에서는 오차가 큰 데이터 샘플에 더 높은 가중치를 부여하여, 다음 트리가 해당 데이터 샘플을 중점적으로 학습하게 한다. 셋째, 정규화 단계에서는 L1 및 L2 정규화를 통해 모델의 복잡도를 제어하고, 이를 통해 과적합을 방지하며 일반화 성능을 높인다. 마지막으로, 병렬 처리 및 트리 Boosting 최적화 단계를 통해 다수의 트리를 병렬로 학습함으로써 기존 Gradient Boosting보다 빠르게 훈련할 수 있다. XGBoost는 결측값 처리에 강력한 성능을 보이며, 이를 통해 데이터 전처리 과정에서 소요되는 시간이 줄어들지만, 비교적 복잡한 구조와 높은 메모리 사용량을 요구하므로 소규모 데이터 세트에서는 과적합 또는 오버헤드가 발생할 수 있다는 단점이 있다.

3.2.4 Extra Trees Regressor

ETR은 RFR과 유사한 결정 트리 기반의 앙상블 학습 알고리즘이다. ETR은 RFR보다 더 많은 무작위성을 도입하여, 학습 과정에서 각 트리의 다양성을 극대화하는 것이 특징이다. ETR은 RFR처럼 여러 개의 결정 트리를 생성하여 예측을 수행하지만, 트리 구성 과정에서 특징 분할과 데이터 표본화에 더 강한 무작위성을 적용하는 것이 차이점이다. 이를 통해 ETR은 학습 속도가 빠르며, RFR보다 과적합을 줄일 수 있는 잠재력을 지닌다. ETR 알고리즘은 다음 네 단계로 이루어진다. 첫째, 데이터 표본화 단계에서는 RFR과 달리 Bootstrap 표본화 과정을 생략하고, 전체 훈련 데이터를 사용하여 의사결정 트리를 학습시킨다. 둘째, 특징 분할 선택 단계에서는 각 노드에서 무작위로 선택된 일부 특징을 기준으로 최적의 분할 대신 임의의 분할 지점을 설정한다. 이러한 무작위성으로 인해 트리 간 상관성이 감소하고, 모델의 일반화 성능이 향상된다. 셋째, 각 트리는 최대 깊이까지 성장하며 가지치기는 수행되지 않는다. 마지막으로, Aggregation 단계에서는 모든 트리의 예측 결과를 평균하여 최종 예측값을 도출한다. ETR 알고리즘은 특징 분할을 무작위로 선택하므로 계산 비용이 감소하고, RFR보다 더 많은 무작위성을 도입하여 모델의 강건성을 높이지만, 데이터가 복잡하고 규칙성이 강한 경우 랜덤한 분할로 인해 모델 성능이 저하될 수 있다.

3.2.5 머신러닝 기반의 회귀 알고리즘의 장단점

본 연구에서 사용된 기계학습 기반의 회귀 알고리즘은 네 가지로 각각의 장단점은 다음과 같이 요약할 수 있다. 첫째, RFR은 비선형 문제에 안정적이고 강력한 성능을 보이지만 다수의 트리를 사용하기 때문에 메모리 사용량이 높은 특징이

있다. 둘째, SVR은 고차원 공간에서 선형적으로 구분되지 않는 데이터를 효과적으로 처리할 수 있지만, 대규모 데이터 세트의 학습에 많은 시간이 소요될 수 있다. 셋째, XGBoost는 빠른 학습속도로 과적합을 방지할 수 있지만, 구조가 복잡해 소규모 데이터 세트에 비효율적일 수 있다. 넷째, ETR은 더 많은 무작위성을 도입하여 빠르게 과적합을 줄이지만, 복잡한 데이터 세트에서는 랜덤한 분할로 인해 성능 저하가 발생할 수 있다.

4. 실험 결과 및 분석

4.1 정량적 분석

정량적 측정 지표로는 x, y 좌표의 경우 평균 제곱근 오차 (Root Mean Square Error, RMSE)를 사용하였고, 층고의 경우 전체 정확도(Accuracy)로 기술하였다. 또한, 각 모델의 장단점을 분석하기 위해 학습 시간과 추론 시간을 같이 측정하여 분석하였다. 실험결과 x, y 좌표를 이용한 RMSE 평가에서는 ETR 알고리즘이 가장 낮은 RMSE를 나타내었고, RFR, XGBoost, SVR 순서의 순위를 나타내었다. WiFi의 신호강도 데이터 세트만을 이용하였지만, 네 가지 학습 모델 모두 2m 이내의 정확도를 나타내었고, 층고 예측 정확도(%)의 경우 ETR과 SVR 알고리즘은 100% 정확도, RFR과 XGBoost는 약 99%를 나타내어 일부 표본의 층고를 제외하고 모두 정확하게 예측함을 확인할 수 있었다. 기계학습 기반의 회귀 알고리즘들의 학습 속도는 Table 2에 표기되어 있으며, XGBoost가 약 87초, SVR이 81초를 나타내었고, RFR과 ETR 알고리즘은 각각 23초와 13초를 나타내었다. 학습 속도의 차이에 더해 전체 테스트 데이터의 추론 시간이 Table 2에 표기되었다. RFR, XGBoost, ETR 알고리즘의 경우 약 0.05초의 추론 시간을 나타내어 서버와의 통신을 통한 실내 측위에 제약사항이 없음을 나타내었지만, SVR의 경우 약 14초를 나타내어 서버 통신 기반의 측위에 SVR 알고리즘은 제약사항이 있음을 확인할 수 있었다. 모든 실험은 파이썬을 이용하여 인텔 Corei9-14900 CPU, 16GB RAM, NVIDIA 4090 GPU 컴퓨터의 우분투 22.04 환경에서 진행되었다.

4.2 시각적 분석

시각적 분석을 위해 테스트 데이터에서 두 궤적의 참조점과 예측점을 Fig. 3과 Fig. 4에 나타내었다. Fig. 3과 Fig. 4의 5-37 및 6-37은 본 연구에서 활용한 데이터 세트에 포함된 궤적정보 ID를 나타낸다. 각각의 알고리즘과 같이 표기된 값은 해당 예측 궤적의 RMSE 값으로, Fig. 3의 궤적에서는 ETR,

RFR, XGBoost, SVR 알고리즘 순으로 RMSE 값이 낮았다. SVR의 경우 건물 복도의 90°이동 변화가 있는 공간에서 예측 궤적이 벗어나는 부분을 확인할 수 있으며, 건물 복도의 끝부분에서 예측점이 참조점과 큰 차이가 있는 것을 확인할 수 있다. Fig. 4에 나타난 궤적의 경우 RFR, ETR, XGBoost, SVR 알고리즘 순으로 RMSE 값이 낮았으며, XGBoost와 SVR의 경우 2m가 넘는 RMSE를 나타내어 실내 공간의 정밀 측위에 다소 미흡한 결과값을 나타내었다. 특히 RFR 알고리즘이 직선 복도의 이동 궤적에 대한 정확한 위치값을 예측한 데 반해 SVR과 XGBoost 알고리즘의 예측점들은 벗어나 있음을 확인할 수 있다.

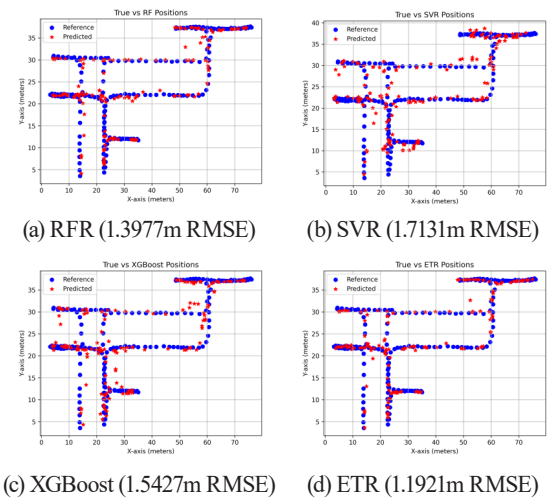


Fig. 3. Visualization of test trajectory-1 (5-37 data)

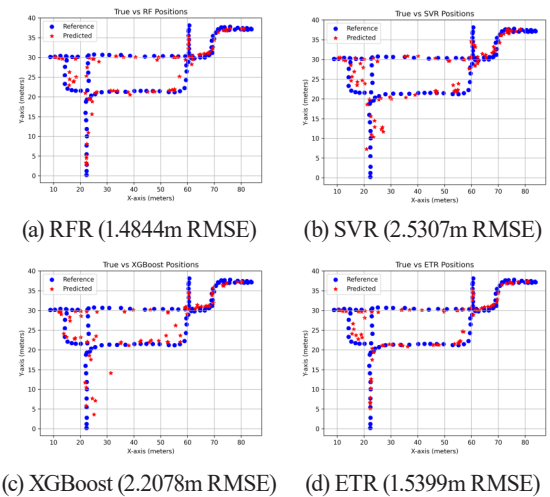


Fig. 4. Visualization of test trajectory-2 (6-37 data)

Table 2. Quantitative performance of regression algorithms

Measures	RFR	SVR	XGBoost	ETR
RMSE (m)	1.48	1.99	1.84	1.27
Accuracy (%)	99.74	100	99.28	100
Training times (sec)	23.27	81.42	87.43	13.34
Inference times (sec)	0.0491	14.33	0.0597	0.0546

5. 요약 및 결론

본 연구에서는 정교하게 구축된 WiFi 신호강도 데이터 세트를 이용하여 다양한 기계학습 기반의 회귀를 적용하여 각각의 예측 성능을 정량적, 시간적으로 분석하였다. 특히, 대부분 연구의 전처리 과정에서 간헐적으로 통신된 AP들을 제외하는 필터링 과정을 적용하는 데 반해, 본 연구에서는 각각의 알고리즘들의 강건성을 실험하기 위해 전체 데이터 세트를 활용한 데 의미를 찾을 수 있다. 이와 같은 접근 전략은 실제 실내 공간에서 자연스럽게 발생할 수 있는 환경을 모사한 것이며, 본 연구에서는 x와 y좌표에 더해 각 건물의 층고를 함께 예측해 분석한 데 추가적인 의미를 부여할 수 있다. 특히 학습 시간과 추론 시간을 함께 분석해 서버와의 통신을 통한 실내 측위 시스템에 적합한 알고리즘을 탐색하였다. 실험 결과 정량적으로 ETR 알고리즘이 약 1.3m의 RMSE 값을 나타내었고 전체 테스트 데이터 세트의 예측에 약 0.05초의 시간을 나타내 실제 실내 측위에 적합함을 확인할 수 있었다. 특히 네 가지 알고리즘 모두 매우 높은 층고 정확도를 나타내어 향후 실내 측위에서는 2차원 위치에 더해 3차원 층고까지 함께 구축하는 것을 고려할 수 있으며, 학습 모델 또한 건물 단위 또는 구획 단위로 구축하여 효율성을 향상할 수 있으리라 판단된다. 향후, 다량의 실내 측위 데이터 세트의 구축이 제한되는 특성을 극복하기 위해 다양한 데이터 증강(augmentation) 알고리즘을 적용하여 실내 측위 연구를 고도화하며 수행할 예정이다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 디지털 국토정보 기술개발사업의 지원으로 수행되었음 (과제번호: RS-2022-00141819).

References

- Alarifi, A., Al-Salman, A., Alsaleh, M., Alnafessah, A., Al-Hadhrami, S., Al-Ammar, M. A., and Al-Khalifa, H. S. (2016), Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors*, Vol. 16, No. 5, 707p.
<https://doi.org/10.3390/s16050707>
- Bianchi, V., Ciampolini, P., and De Munari, I. (2018), RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes. *IEEE Transactions on Instrumentation and Measurement*, Vol. 68, No. 2, pp. 566-575.
<https://doi.org/10.1109/TIM.2018.2851675>
- Grinsztajn, L., Oyallon, E., and Varoquaux, G. (2022), Why do tree-based models still outperform deep learning on typical tabular data?, *Advances in neural information processing systems*, Vol. 35, pp. 507-520.
<https://doi.org/10.48550/arXiv.2207.08815>
- Gu, Y., Lo, A., and Niemegeers, I. (2009), A survey of indoor positioning systems for wireless personal networks. *IEEE Communications surveys & tutorials*, Vol. 11, No. 1, pp. 13-32.
<https://doi.org/10.1109/SURV.2009.090103>
- He, S. and Chan, S.H.G. (2015), Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys & Tutorials*, Vol. 18, No. 1, pp. 466-490.
<https://doi.org/10.1109/COMST.2015.2464084>
- Ji, W., Zhao, K., Zheng, Z., Yu, C., and Huang, S. (2021), Multivariable fingerprints with random forest variable selection for indoor positioning system. *IEEE Sensors Journal*, Vol. 22, No. 6, pp. 5398-5406.
<https://doi.org/10.1109/JSEN.2021.3103863>
- Jiang, W., Cao, Z., Cai, B., Li, B., and Wang, J. (2021), Indoor and outdoor seamless positioning method using UWB enhanced multi-sensor tightly-coupled integration. *IEEE Transactions on Vehicular Technology*, Vol. 70, No. 10, pp. 10633-10645.
<https://doi.org/10.1109/TVT.2021.3110325>
- Jondhale, S.R., Mohan, V., Sharma, B.B., Lloret, J., and Athawale, S.V. (2022), Support vector regression for

- mobile target localization in indoor environments. *Sensors*, Vol. 22, No. 1, 358p.
<https://doi.org/10.3390/s22010358>
- Khassanov, Y., Nurpeiissov, M., Sarkytbayev, A., Kuzdeuov, A., and Varol, H.A. (2021), Finer-level sequential wifi-based indoor localization. In *2021 IEEE/SICE International Symposium on System Integration (SII)* pp. 163-169.
<https://doi.org/10.1109/IEEECONF49454.2021.9382623>
- Kushki, A., Plataniotis, K.N., and Venetsanopoulos, A.N. (2007), Kernel-based positioning in wireless local area networks. *IEEE transactions on mobile computing*, Vol. 6, No. 6, pp. 689-705.
<https://doi.org/10.1109/TMC.2007.1017>
- Liu, H., Darabi, H., Banerjee, P., and Liu, J. (2007), Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 37, No. 6, pp. 1067-1080.
<https://doi.org/10.1109/TSMCC.2007.905750>
- Liu, M., Liao, X., Zhang, Y., and Gao, Z. (2024), A Multi-scale Spatial-temporal Features Fusion Framework for Indoor Localization. *IEEE Sensors Journal*, Vol. 24, No. 15, pp. 23098-23107.
<https://doi.org/10.1109/JSEN.2024.3395772>
- Nessa, A., Adhikari, B., Hussain, F., and Fernando, X.N. (2020). A survey of machine learning for indoor positioning. *IEEE access*, Vol. 8, pp. 214945-214965
<https://doi.org/10.1109/ACCESS.2020.3039271>
- Nosrati, L., Fazel, M.S., and Ghavami, M. (2022), Improving indoor localization using mobile UWB sensor and deep neural networks. *IEEE Access*, Vol. 10, pp. 20420-20431.
<https://doi.org/10.1109/ACCESS.2022.3151436>
- Shang, F., Su, W., Wang, Q., Gao, H., and Fu, Q. (2014), A location estimation algorithm based on RSSI vector similarity degree, *International Journal of Distributed Sensor Networks*, Vol. 10, No. 8, 371350p.
<https://doi.org/10.1155/2014/371350>
- Shwartz-Ziv, R. and Armon, A. (2022), Tabular data: Deep learning is not all you need. *Information Fusion*, Vol. 81, pp. 84-90.
<https://doi.org/10.1016/j.inffus.2021.11.011>
- Singh, N., Choe, S., Punmiya, R., and Kaur, N. (2022), Xgbloc: Xgboost-based indoor localization in multi-building multi-floor environments. *Sensors*, Vol. 22, No.17, 6629p.
<https://doi.org/10.3390/s22176629>
- Zafari, F., Gkelias, A., and Leung, K.K. (2019), A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, Vol. 21, No. 3, pp. 2568-2599.
<https://doi.org/10.1109/COMST.2019.2911558>