

XGboost 기반의 WiFi 신호를 이용한 실내 측위 기법

황치곤¹ · 윤창표² · 김대진^{3*}

Indoor positioning method using WiFi signal based on XGboost

Chi-Gon Hwang¹ · Chang-Pyo Yoon² · Dae-Jin Kim^{3*}

¹Invited Professor, Dept. of Computer Engineering, IIT, Kwangwoon University, Seoul, 01897 Korea

²Assistant Professor, Dept. Of Computer & Mobile Convergence, GyeongGi University of Science and Technology, Siheung-si, 15073 Korea

^{3*}Assistant Professor, Institute for Image & Cultural Contents, Dongguk University, Seoul, 04626 Korea

요약

위치를 정확하게 측정하는 것은 다양한 서비스를 제공하는 데 필요하다. 실내 측위를 위한 데이터는 스마트폰의 앱을 통해 WiFi 장치로부터 RSSI 값을 측정한다. 이렇게 측정된 데이터는 기계학습의 원시 데이터가 된다. 특징 데이터는 측정된 RSSI 값이고, 레이블은 측정된 위치에 대한 공간의 이름으로 한다. 이를 위한 기계학습 기법은 분류에 효율적인 기법을 적용하여 WiFi 신호만으로 정확한 위치를 예측하는 기법을 연구하고자 한다. 앙상블은 하나의 모델보다 다양한 모델을 통하여 더 정확한 예측값을 구하는 기법으로, bagging과 boosting이 있다. 이 중 Boosting은 샘플링한 데이터를 바탕으로 모델링한 결과를 통해 모델의 가중치를 조정하는 기법으로, 다양한 알고리즘이 있다. 본 연구는 위 기법 중 XGboost를 이용하고, 다른 앙상블 기법과 이용한 수행결과를 바탕으로 성능을 평가한다.

ABSTRACT

Accurately measuring location is necessary to provide a variety of services. The data for indoor positioning measures the RSSI values from the WiFi device through an application of a smartphone. The measured data becomes the raw data of machine learning. The feature data is the measured RSSI value, and the label is the name of the space for the measured position. For this purpose, the machine learning technique is to study a technique that predicts the exact location only with the WiFi signal by applying an efficient technique to classification. Ensemble is a technique for obtaining more accurate predictions through various models than one model, including bagging and boosting. Among them, Boosting is a technique for adjusting the weight of a model through a modeling result based on sampled data, and there are various algorithms. This study uses Xgboost among the above techniques and evaluates performance with other ensemble techniques.

키워드 : 실내 측위, RSSI(Receiver Signal Strength Indicator), Bagging, Boosting, XGboost

Keywords : Indoor positioning, RSSI(receiver signal strength indicator), Bagging, Boosting, XGboost

Received 29 November 2021, Revised 9 December 2021, Accepted 19 December 2021

* Corresponding Author Dae-Jin Kim(E-mail:sampoo00@dongguk.edu, Tel:+82-2-2264-0160)

Assistant Professor, Institute for Image & Cultural Contents, Dongguk University, Seoul, 04626 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.1.70>

print ISSN: 2234-4772 online ISSN: 2288-4165

© 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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

이동통신의 발전과 위치를 기반으로 하는 서비스들이 증가함에 따라 위치의 정확한 산출은 중요한 이슈이다. 위치를 산출하기 위해 사용하는 GIS(Geographic Information System) 서비스와 이것을 이용하는 실외 서비스는 위성을 이용하는 방법이 주로 이루고 있어 실내 측위에는 부족하다. 실내 위치를 측정하기 위한 기술로는 WiFi, BLE(Bluetooth Low Energy) beacon 등을 이용하는 연구가 있다[1]. 실내의 위치 측정 기술로 WiFi 신호의 세기를 이용하여 측정하거나, BLE Beacon을 통해 가까운 거리에서 발생한 Beacon 신호를 이동통신 기기를 이용하여 수신하여 서비스에 이용하는 것으로 원격 제어하거나 위치 기반의 서비스를 하는 등 많은 분야에 적용할 수 있다. 그러나 WiFi, BLE beacon 신호를 바탕으로 한 위치 인식은 여러 가지 문제가 있다. 첫째, 인식 위치의 환경에 따라 위치의 오차범위가 달라질 수 있다. 둘째, 주변 장애물이 신호의 왜곡을 일으키기 때문에 정확도가 떨어뜨릴 수 있다. 셋째, 인접한 장소에 있는 WiFi 신호가 서로 뒤바뀌는 현상이 발생할 수 있다 [2][3]. 이에 다양한 위치에 대한 WiFi 신호를 바탕으로 정확한 측위를 할 수 있는 기법이 필요로 하다.

최근 기계학습을 통한 분류 기법들이 연구되고 있다. 분류 기법 중 의사결정 트리(decision tree) 기법과 앙상블 기법이 있다. 의사결정 트리는 정확도는 높으나 과적합(over fitting)이 발생할 경우가 많다. 앙상블 기법은 하나의 모델보다 다양한 모델을 통하여 더 정확한 예측 값을 구하는 기법으로 의사결정 트리의 단점을 보완할 수 있으며, bagging과 boosting이 있다. bagging은 의사결정 트리의 과적합의 문제를 해결하기 위해 여러 번의 샘플링과 모델을 학습한 결과를 집계한 방식이고, boosting은 샘플링한 모델의 오분류에 대한 가중치를 계산하여 다음 샘플링의 학습에 적용한 기법이다 [4]. 그리고 boosting에는 AdaBoost, GradientBoost, XGBoost, CatBoost와 같은 기법이 있다.

본 논문에서는 실내 측위를 위하여 부스팅 기법 중 XGboost 기법을 적용하고, 다른 기법들과 비교 평가하였다. 본 논문은 2장에서 실내 측위와 기계학습 기법에 관해 관련 연구를 기술하고, 3장에서 측정된 데이터의 전처리와 예측을 위한 XGBoost 기법을 적용한다. 4장 다른 기법과 비교 평가를 적용하고, 5장은 본 논문의 결

론을 기술한다.

II. 관련 연구

WiFi 장치의 RSSI는 비용이 저렴하고, 측정하기 쉬우며, 정밀도가 높다는 이유로 실내 측위 방식으로 많이 사용한다. 최근 RSSI 신호를 기반으로 사용하는 경우 기존 실내의 측위 방식보다 정밀도가 향상되고 있기 때문이다 [5].

이 WiFi의 신호는 오프라인 단계에서 실내 측위는 높은 정확도를 보였지만, 온라인 단계에서 사용자의 위치를 계산하는데 많은 시간이 소요되는 문제가 발생한다 [6]. 그리고 기계학습에 필요한 무선 신호 데이터의 수집 과정에서 왜곡되거나 학습에 적합하지 않은 데이터가 포함되어 위치 인식의 정확도가 낮아지는 결과가 발생한다. 신호 테이블에 포함된 특정 위치의 신호 값 중에는 학습에 적절하지 않은 왜곡되거나 일시적으로 출현한 신호와 다른 위치와 중복되지만, 효율이 떨어지는 미약한 신호 값이 포함된다[7]. 위와 같은 문제로 위치를 특정할 수 없으며, 위치를 측정하는 사람이 장소로 구분되는 블록의 경계에 위치하였을 때 어떠한 장소에 속하는 위치인지를 결정하는데 문제가 발생한다. 즉, 수집된 데이터는 장소로 구분하여 라벨링 되고 특정된 위치와 그렇지 않은 위치에 대한 경계가 모호하게 되어 라벨링들 사이에 존재하는 중간 위치의 경우 해당 위치가 어느 라벨에 속하는지 결정하는 과정에서 판정이 잘못되는 문제가 발생한다. 이처럼 라벨링 되어 특정된 위치와 그렇지 않은 주변 위치에 중복으로 나타나지만, 그 신호가 미약하거나 지속적이지 않은 데이터를 포함한다면 위치의 판단은 매우 어렵게 된다.

WiFi 장치에서 RSSI가 가리키는 무선 네트워크 식별자인 SSID(Service Set Identifier)를 통해 결정 트리를 생성하고 기계학습의 일종인 Random Forest를 활용하여 의사결정 트리를 취합하는 시스템을 구현하여 실행 속도와 정확도를 높이는 연구 또한 진행되었다[8].

의사결정 트리를 사용한 분석 기법은 공통으로 데이터 과적합의 문제가 있고, Random Forest는 과적합을 줄이기 위해서는 더 많은 양의 트리를 만든다[9]. 이와 같이 의사결정 트리는 과적합으로 인해 정확도가 낮아지는 성능의 문제가 있다. 이를 해결하기 위한 기법으로

양상블 기법이 제시되고 있다. 양상블을 이용하는 기법에는 bagging과 boosting 기법이 있으며, boosting은 여러 개의 샘플링을 이용하여 각 샘플을 순차적으로 학습을 수행하는 기법으로, 이전 샘플의 학습 결과에 대한 손실을 다음 샘플의 학습에서 사용할 가중치 조정으로 샘플들의 학습을 진행하여 샘플의 학습에 대한 정확률을 높이는 방법이다[4].

최근 boosting 기법 중 XGboost 알고리즘이 효율이 높다. XGBoost는 지도 학습에 속하는 부스팅 알고리즘으로, 그래디언트 부스트 트리(gradient boosted tree)를 기반으로 하는 양상블 알고리즘입니다 [10][11]. 그림 1은 XGboost 기법에서 샘플이 순차적으로 가중치를 갱신하며 최종 모델(model t)로 수행되는 과정이다.

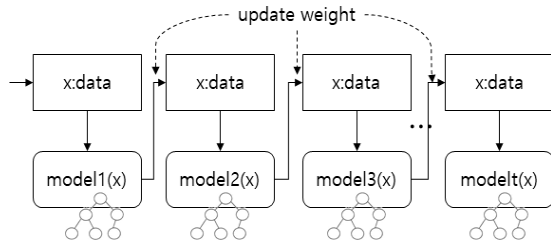


Fig. 1 Model workflow in XGboosting

그림 1은 순차적인 학습 과정을 통해 이전 모델이 달성한 결과에 따른 오차를 이용하여, 다음 학습 모델의 가중치를 갱신하여 예측의 정확률을 높인다. 또한, 정규화 항을 통해 random Forest 기법의 과적합 문제를 해결하는 기법으로, 수식 (1)은 모델 생성을 위한 기본 수식이다.

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_i(x_i) \quad (1)$$

수식 (1)에서 t는 샘플링된 모델의 총수로 $\hat{y}_i^{(t)}$ 는 XGboost의 최종 트리 모델이고, $\hat{y}_i^{(t-1)}$ 는 이전 샘플에 관한 결과이다. 이전 샘플의 결과 현재 샘플의 모델인 $f_i(x_i)$ 를 갱신시켜 새로운 모델의 최종 모델을 생성한다. 목적함수는 최적의 분류기를 찾기 위해 손실함수 L을 적용하여 최적의 알고리즘을 찾는다. 이에 대한 수식이 수식 (2)와 같다.

$$Obj^{(t)} = \sum_{i=1}^t L(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_i) \quad (2)$$

수식 (2)에서 y_i 는 실제 값이고, $\hat{y}_i^{(t)}$ 는 예측된 값이며, $L(y_i, \hat{y}_i^{(t)})$ 는 손실함수이고, $\Omega(f_i)$ 는 정규화이다.

XGBoost는 많은 분야에 대한 확장성과 뛰어난 정확성이 장점이다. XGBoost의 확장성은 병렬 및 분산 컴퓨팅을 사용하여 알고리즘이 가중치를 사용하여 최소 데이터를 처리하는 방식으로 GBM(Gradient Boosting Method)과 통합함으로써 사용 가능한 모든 하드웨어 리소스를 활용하여 예측 성능과 처리 속도를 극대화할 수 있다[12].

따라서 본 논문에서는 XGboost 기법을 적용하여 여러 공간의 WiFi 장치에서 발생하는 RSSI 신호 세기를 측정하여 실내 공간의 측위하는 방안을 제안한다.

III. 제안된 기법 구현 및 평가

3.1. 데이터 수집 및 전처리

테스트에 사용되는 공간에 WiFi 공유기에서 발생하는 신호를 측정하기 위해 스마트폰 애플리케이션을 이용하여 데이터를 수집하고, 측정된 위치에서 발생하는 모든 WiFi 신호를 학습을 위한 특징 데이터가 되고, SSID를 기반으로 레이블을 부여한다.

테스트에 사용되는 데이터는 3개의 공간에 한해서 각 영역에서 수신되는 WiFi 공유기의 신호를 6,000건씩 수집하여 전체 18,000건의 데이터를 수집하였다. 3개의 공간은 레이블로 사용될 이름으로써 A 영역[320], B 영역[324], C 영역[326]으로 하고, 수집된 데이터의 특징 데이터는 -85 dbm 이하의 신호인 경우와 수신되지 않은 WiFi 공유기의 신호와 레이블에 대한 one-hot 인코딩을 전처리 작업을 통해서 수행한다. 이에 따라 수행한 원시 데이터(raw data)를 전처리한 결과는 표 1과 같다.

Table. 1 Sample of raw data with preprocessing

label	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
100	-73	1	-72	1	-69	1	-77	-73	-37	-36
100	-73	1	-72	1	-69	1	-77	-73	-37	-36
100	-73	1	-72	1	-69	1	-77	-73	-37	-36
100	-73	1	-72	1	-69	1	-77	-73	-37	-36
100	-73	1	-72	1	-69	1	-77	-73	-37	-36
010	-43	-69	-58	-84	-46	-84	-43	-42	-70	-70
010	-43	-71	-57	-85	-46	-84	-43	-43	-65	-69

label	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
010	-45	-68	-57	-84	-46	-82	-43	-44	-69	-66
010	-43	-71	-58	-84	-46	-84	-43	-43	-66	1
010	-41	-61	-63	-85	-43	-83	-40	-42	-73	-71
001	-64	-82	-85	-55	-65	-77	-76	-65	1	1
001	-66	-83	1	-55	-65	-72	-80	-65	1	1
001	-63	-81	-85	-55	-66	-74	-78	-64	1	1
001	-66	-85	-84	-54	-65	-76	-73	-67	-85	-85
001	-62	1	1	-63	-64	-79	-77	-61	-85	1

3.2. 학습 과정 및 구현

학습에 사용된 하이퍼 파라미터는 표 2와 같다. XGboost에서 사용하는 하이퍼 파라미터는 86개의 파라미터가 있지만, 학습에 많은 영향을 미치는 것들만 표 2에서 기술하였고, 여러 회차에 따른 학습을 통해 가장 효율이 높게 나오는 값들을 선택하였다.

Table. 2 Hyper-parameters used for training

hyper parameter	value
learning_rate	0.01
num_boost_rounds	500
min_child_weight	1
max_depth	6
subsample	0.2
early_stopping_rounds	100

```
>> confusion_matrix
[[1196  0  10]
 [  40 1138  13]
 [  10  7 1186]]
```

Fig. 2 Confusion matrix for experimental results

입력 데이터에 대한 학습 데이터와 테스트 데이터에 대한 비율은 18,000건이라는 원본데이터의 부족으로 인해 학습을 위한 비율을 높여 좀 더 나은 모델을 생성하기 위하여 8:2로 하였다. 그 결과로 산출된 혼동행렬 (Confusion Matrix)은 그림 2와 같다.

혼동행렬은 모델의 분류 결과에 대한 지표로 성능을 평가할 수 있다. 테스트 데이터에 대한 예측된 분류 결과와 실제 레이블에 대응되는 건수를 나타낸 것으로 대각선상의 건수가 정확하게 예측된 것이다. 이를 바탕으로 정밀도, 재현율, f1-score를 산출한다. 모델 평가에 대한 보고서는 그림 3과 같고, 각 영역에 대한 정밀도는

96%에서 99% 사이의 값이 나왔고, 재현율은 96%, 99%로 나왔으며, f1-score는 97%, 98%로 높은 결과를 산출하였다. 이에 따른 전체적인 정확률은 98%로 나왔다.

```
>> model evaluation report <<
      precision    recall  f1-score   support

   320     0.96     0.99     0.98     1206
   324     0.99     0.96     0.97     1191
   326     0.98     0.99     0.98     1203

 accuracy                   0.98     3600
```

Fig. 3 Model evaluation report

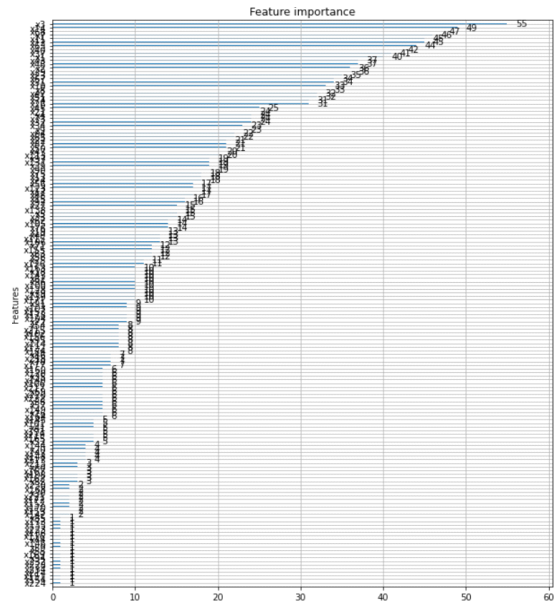


Fig. 4 Feature importance according to experimental results

그림 4는 학습에서 특징 데이터들에 대한 중요도 순서이다. 현재 실험을 위하여 사용된 데이터는 3개의 영역에 대한 데이터들 수집하여 측정된 것이기 때문에 많은 영향을 주는 특징을 가지는 WiFi 공유기는 상위 일부의 데이터임을 확인할 수 있다. 그러나 영역이 확장되면 특징의 중요도는 달라질 것이다.

IV. 비교 평가

양상블 기법은 bagging과 boosting 기법이 있다. 본

논문에서는 boosting 기법 중 XGboosting 기법을 이용하여 실험한 결과를 제시하였다. 이에 본 장에서는 bagging 기법의 Random Forest와 비교를 통하여 평가한다. 테스트에 사용한 데이터, 하이퍼 파라미터, 사용 환경은 3장의 실험과 같다. 이 결과는 표 3과 같다.

Table. 3 Comparison of performance evaluation of Random Forest and XGboost

algorithm measure	Random Forest	XGboost
accuracy(%)	94.27	98.13
run time(sec.)	1.5806	5.1938

표 3은 정확률과 실행 시간을 평균적으로 비교한 것이다. 매번 실행할 때마다 샘플링의 차이에 따라 혼동행렬과 정확률이 조금씩 달라지므로 한 번의 실행결과로는 비교할 수 없으므로 여러 번 수행한 평균으로 비교하였다. 이에 XGboost 기법이 정확률이 평균적으로 약 3.86% 높게 나왔으나 사실상 90% 이상의 정확률에서 1%의 차이도 많은 차이이므로 높은 성능 차이라고 볼 수 있고, 실행 시간에서 느린 것은 bagging과 boosting의 기법의 특징 차이로 보이며, 병렬 처리로 수행하면 그 차이는 감소할 것으로 보인다.

V. 결 론

실내 측위에는 다양한 기법이 있다. 이에 최근의 기계 학습 기법을 통해 관련 연구에서 제시한 실내 측위의 단점을 해결하고자 하였다. 본 논문에서는 실내 측위를 위한 기법으로 XGboosting 기법을 적용하여 스마트 폰을 통해 수집한 실제 데이터를 기반으로 실험을 수행하였고, 수행된 결과에 따라 지역을 구분하는 분류 기법으로 적합함을 보였다. 향후 연구는 본 연구를 바탕으로 실내 경로 예측 및 이동 경로 안내를 위한 기법으로 확장할 것이다.

References

[1] K. Konstantinos and T. Orphanoudakis, "Bluetooth beacon based accurate indoor positioning using machine learning,"

in *2019 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference*, pp. 1-6, Sep. 2019.

[2] D. J. Kim, C. G. Hwang, and C. P. Yoon, "Learning data preprocessing technique for improving indoor positioning performance based on machine learning," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 24, no. 11, pp. 1528-1533, Nov. 2020.

[3] C. G. Hwang, C. P. Yoon, and D. J. Kim, "Indoor positioning system using Xgboosting," *Proceedings of the Korean Institute of Information and Commucation Sciences Conference*, vol. 45, pp. 492-494, 2021

[4] S. González, S. García, J. Del Ser, L. Rokach, and F. Herrera, "A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities," *Information Fusion*, vol. 64, pp. 205-237, Dec. 2020.

[5] S. H. Oh and J. G. Kim, "WiFi Positioning Based on PSO in 3GPP Indoor Environments," *The Journal of Korean Institute of Communications and Information Sciences*, vol. 46, no. 9, pp. 1440-1448, Sep. 2021.

[6] D. B. Ninh J. He, V. T. Trung, and D. P. Huy, "An effective random statistical method for indoor positioning system using WiFi fingerprinting," *Future Generation Comput. Syst.*, vol. 109, pp. 238-248, Aug. 2020.

[7] H. G. Shin, Y. H. Choi, and C. P. Yoon, "Movement Path Data Generation from Wi-Fi Fingerprints for Recurrent Neural Networks," *Sensors*, vol. 21, no. 8, pp. 2823, Apr. 2021.

[8] S. Lee, J. Kim, and N. Moon, "Random forest and WiFi fingerprint-based indoor location recognition system using smartwatch," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, pp. 6, Feb. 2019.

[9] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, Oct. 2001.

[10] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp. 785-794, 2016.

[11] H. Mo, H. Sun, J. Liu, and S. Wei, "Developing window behavior models for residential buildings using XGBoost algorithm," *Energy and Buildings*, vol. 205, no. 15, pp. 109564, Dec. 2019.

[12] K. K. Yun, S. W. Yoon, and D. Won, "Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering proces," *Expert Systems with Applications*, vol. 186, pp. 115716, Dec. 2021.



황치곤(Chi-Gon Hwang)

2012년 광운대학교 컴퓨터과학과 (공학박사)
2006년~2015년:(주)인찬 연구원
2016년~2018년: 경민대학교 인터넷정보과 교수
2019년~현재: 광운대학교 정보과학교육원 컴퓨터공학과 교수
※관심분야 : 모바일 클라우드, 온톨로지, 기계학습, NLP



윤창표(Chang-Pyo Yoon)

1998년 광운대학교 전자계산학과 (이학사)
2001년 광운대학교 컴퓨터과학과 (공학석사)
2012년 광운대학교 컴퓨터과학과 (공학박사)
2012년~현재: 경기과학기술대학교 컴퓨터모바일융합과 교수
※관심분야 : 기계학습, 모바일 시스템, 네트워크 보안, 무선 네트워크



김대진(Dae-Jin Kim)

1998년 대진대학교 전자공학과 (공학사)
2000년 대진대학교 전자공학과 (공학석사)
2010년 대진대학교 전기전자통신공학과 (공학박사)
2017년~현재: 동국대학교 영상문화콘텐츠연구원 교수
※관심분야 : Deep Learning, Object Detection, Action Recognition, Face Recognition, CODEC, Watermark, Fingerprint