

Xgboosting 기법을 이용한 실내 위치 측위 기법

황치곤^{1*} · 윤창표² · 김대진^{3,*}

¹광운대학교 · ²경기과학기술대학교 · ³동국대학교

Indoor positioning system using Xgboosting

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

¹Kwangwoon University · ²GyeongGi University of Science and Technology · ³Dongguk University

E-mail : duck1052@kw.ac.kr / cpyoon@gtec.ac.kr / sampoo00@dongguk.edu

요 약

기계학습에서 분류를 위한 기법으로 의사결정트리 기법을 이용한다. 그러나 의사결정트리는 과적합의 문제로 성능이 저하되는 문제가 있다. 이러한 문제를 해결하기 위해 여러 개의 부트스트랩을 생성하여 각 자료를 모델링하여 학습하는 Bagging 기법, 샘플링한 데이터를 모델링하여 가중치를 조정하여 과적합을 감소시키는 Boosting과 같은 기법으로 이를 해결할 수 있다. 또한, 최근에 Xgboost 기법이 등장하였다. 이에 본 논문에서는 실내 측위를 위한 wifi 신호 데이터를 수집하여 기존 방식과 Xgboost에 적용하고, 이를 통한 성능평가를 수행한다.

ABSTRACT

The decision tree technique is used as a classification technique in machine learning. However, the decision tree has a problem of consuming a lot of speed or resources due to the problem of overfitting. To solve this problem, there are bagging and boosting techniques. Bagging creates multiple samplings and models them using them, and boosting models the sampled data and adjusts weights to reduce overfitting. In addition, recently, techniques Xgboost have been introduced to improve performance. Therefore, in this paper, we collect wifi signal data for indoor positioning, apply it to the existing method and Xgboost, and perform performance evaluation through it.

키워드

Decision Tree, Bagging, Boosting, Random Forest, Xgboost (Extreme Gradient Boosting)

1. 서 론

최근 기계학습을 통한 분류기법들이 지속적으로 나오고 있다. 대표적으로 의사결정트리이며, 이 의사결정 트리 기법은 Xgboost는 순차적인 형태로 의사결정 트리를 생성하는 최적화된 그래디언트 트리 부스팅 시스템이다[1]. Xgboost는 모든 컴퓨팅 환경에서 빠르게 수행할 수 있어 새로운 속성을 모델링하고 레이블을 분류하는 데 높은 성능을 가지고 있다[2]. Xgboost 알고리즘의 적용은 표와 같이 구조화된 데이터 셋으로 구현될 때 높은 성능을 얻을 수 있다. Xgboost 알고리즘은 의사결정에 대한 솔루션의 특정 조건에 따라 계산되는 의사

결정 트리를 기반으로 한 접근 방식으로 시작되었다. 의사 결정 트리의 과적합의 문제를 해결하기 위해 배깅(Bagging)이라는 기법이 있고, 이에 대한 오분류에 대한 가중치를 계산하여 다음 샘플링의 학습에 적용한 부스팅(Boosting) 기법이 있다[1][3]. 이들 중 부스팅 기법 중 Xgboost 기법이 가장 대표적으로 이용되고 있다. 이 기법을 실내 측위 데이터에 적용하여 성능을 평가하였다.

이동통신의 발전에 따라 위치를 기반으로 하는 서비스들이 증가함에 따라 위치의 정확한 산출은 중요한 이슈이다. GIS (Geographic Information System) 서비스와 GPS (Global Positioning System)를 이용하는 실외 서비스의 경우는 위성을 이용하는 방법 등은 실내의 위치를 측정하기에는 부족함이 있고, 실내 위치를 측정하기 위한 기술로는

* corresponding author

WiFi, BLE (Bluetooth Low Energy) beacon 등을 이용하는 연구가 지속되고 있다[4].

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 적용시킬 수 있다[5].

본 논문에서는 실내 측위를 위하여 Xgboost 기술을 적용하여 정확성과 성능향상을 위한 연구를 수행한다. 이에 따라 본 논문은 2장에서 관련 연구를 기술하고, 3장에서 본문, 4장에서 결론을 기술한다.

II. 관련연구

Xgboost는 지도 학습(supervised learning)에 속하는 부스팅 알고리즘으로, 그래디언트 부스트 트리를 기반으로 하는 앙상블 알고리즘이다[1]. 직렬 훈련 프로세스를 통해 "강한" 분류기를 달성하기 위해 "약한" 분류기의 예측을 통합한다. 이에 정규화 항을 추가하여 과적합의 문제를 해결할 수 있다. 병렬 및 분산 컴퓨팅은 학습 과정을 빠르게 만들어 모델링의 향상할 수 있다. Xgboost의 계산 과정은 수식1에 따라 수행된다.

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

$\hat{y}_i^{(t)}$ 는 최종 모델이다. $\hat{y}_i^{(t-1)}$ 는 이전에 생성된 모델이고, $f_t(x_i)$ 는 새로 생성된 모델이며, t는 기본 모델의 수이다. 즉, 최종 모델은 이전에 생성된 모델에 새로 생성된 모델을 결합함으로써 생성된다 [1-3]. 이 Xgboost는 모든 시나리오에서 확장성과 뛰어난 정확성을 가진다. 본 논문은 위의 기법을 실내 측위 데이터를 수집하여 학습 데이터로 적용한다.

III. 제안 기법 및 성능 평가

제안하는 기법은 실내 측위 데이터를 수집하여 Xgboosting 기법으로 실내에서 위치를 지정한다. 실내 데이터의 수집은 특정 지역에 여러 대의 wifi 장치를 통해서 여러 위치에서 데이터를 수집하고, 이를 전처리를 수행한 데이터를 학습 데이터로 사용하고 수집된 각 위치를 레이블로 사용한다. 학습 데이터와 테스트 데이터는 기본적인 비율인 6:4의 비율로 구분한다. 이에 따른 수집된 데이터는 표 1과 같다.

표 1. 추출된 데이터의 학습데이터 샘플

x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	...	y
-77	-70	-76	-61	-75	-75	-87	-89	...	326
-74	1	-77	-61	-74	-76	-89	-88	...	326
-73	-69	-72	1	-72	-71	1	-87	...	326
-78	-74	-65	1	-75	-63	-76	-88	...	326
...	326
-80	-75	-74	1	-77	-72	-83	-89	...	326
1	1	1	1	-63	-91	1	1	...	324
1	1	1	1	-64	1	1	1	...	324
1	-91	1	1	1	-87	1	1	...	324
1	-93	1	1	-62	1	1	1	...	324
...	324
1	-86	1	-79	-61	1	1	1	...	324
-83	1	1	1	1	1	1	1	...	116
1	1	-86	1	1	-86	-82	1	...	116
-82	1	1	1	1	1	1	1	...	116
-83	1	-85	1	1	1	1	1	...	116
...	116
-82	1	-84	1	1	-85	-81	1	...	116

표1은 약 500개의 와이파이 장치로부터 수신된 데이터를 샘플링한 것으로 x₁~x_n은 와이파이 장치이며 각 신호를 측정된 위치가 y로 레이블에 해당한다. 수신된 값은 RSSI값으로 신호의 강도를 의미하며 수신되지 않거나 무의미한 값(null)은 데이터 처리에 영향을 미치지 않도록 1의 값으로 전처리하였다. 이를 Xgboost 알고리즘에 따라 처리한 실제 측위된 레이블과 예측값을 비교하였다. Xgboost 알고리즘과 RandomForest 기법으로 위 데이터로 동일한 하이퍼파라미터를 이용하여 성능평가를 수행한 결과는 표2와 같다.

표 1. 알고리즘에 따른 성능평가

measure \ algorithm	RandomForest	Xgboost
accuracy(%)	80.02	92.4
run time(sec.)	1.306833	0.555856

IV. 결 론

실내 측위를 위한 기법으로 wifi신호를 기반으로 하는 fingerprint 방식, 삼각 측량법 등과 같은 방식을 사용하고, 최근에 머신러닝 기법을 적용하여 측위를 하는 기법들이 출현하였다. 이에 본 논문에서는 분류를 위한 기계학습 기법 중 Xgboost 기법이 높은 성능을 제공하고 있어 이를 통해 측위를 위해 wifi 신호를 바탕으로 하여 적용하는 방안을 제시하였다. 이는 앞으로 wifi신호에 대한 정제와 다양한 의사결정트리 기법에 적용하여 정확도와 속도를 향상시킬 수 있도록 하여 일반화 시켜야 한다.

References

- [1] 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.
- [2] S. Bhattacharya, S. R. Krishnan, P. K. R. Maddikunta, R. Kaluri, S. Singh, T. R. Gadekallu, M. Alazab, and U. Tariq, "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU," *Electronics*, Vol. 9, No. 2, pp. 2-17. 2020.
- [3] 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, 115716, 2021.
- [4] J. Kim and B. G. Gu, "Machine Learning Method based on Beacon Signal Strength Pattern for Deciding Indoor Presence of Use," *Journal of Korean Institute of Information Technology*, vol. 18, no. 8, pp. 1-8, Aug. 2020.
- [5] 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, 2020.