논문 18-43-11-27 The Journal of Korean Institute of Communications and Information Sciences '18-11 Vol.43 No.11
https://doi.org/10.7840/kics.2018.43.11.1976

랜덤 포레스트 분류기를 이용한 IoT 지원 지문 실내 위치인식

위 예 교•, 이 상 문*, 황 승 훈°

IoT-Aided Wi-Fi Based Fingerprint Indoor Positioning Using Random Forest Classifier

Yiqiao Wei•, Sang-Moon Lee*, Seung-Hoon Hwang°

요 약

Wi-Fi 기반 실내 위치인식 기술은 가장 인기 있는 실내 위치인식 기술 중 하나이다. 본 연구에서는 랜덤포레스트 분류기를 이용한 지문 실내 위치인식 시스템을 제안한다. 지문 데이터베이스는 IoT 장치와 개발 프로그램으로 구성된다. 그리고 데이터베이스는 74개의 목표 위치가 실제 실내 환경에서 사용자 위치를 예측할 수 있도록 기계 학습 분류기를 학습시킨다. 시뮬레이션 결과에 따르면 랜덤포레스트 분류기가 KNN 분류기 및 SVM 분류기보다 우수한 최대 94%의 위치결정 정밀도를 보입니다. 실시간 실험은 랜덤포레스트 분류기를 적용한 시스템이 91% 성공률로 4m의 정밀 실내 배치를 달성할 수 있음을 확인하였다.

Key Words : Fingerprint Indoor positioning, IoT, Received Signal Strength (RSS), Random Forest.

ABSTRACT

Wi-Fi based fingerprint indoor positioning technology is known as one of the most popular indoor positioning technologies. In this work, an internet of things (IoT) aided fingerprint indoor positioning system using Random Forest classifier has been proposed. The fingerprint database is constructed with IoT device and developed program. Then database is used to train machine learning classifier to be able to predict user position in a real indoor environment with 74 target locations. The simulation results show that Random Forest classifier is more powerful than KNN classifier and SVM classifier with positioning accuracy up to 94%. The real-time experiment verified that Random Forest classifier applied system can achieve 4 meters precision indoor positioning with 91% success rate.

Ⅰ. Introduction

Nowadays, Location Based Services (LBSs) have attracted lots of attention and this greatly drives the development of indoor positioning technologies. And more and more practical IoT applications are

※ This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program-Industrial site core technology Early development project) (10084909, Development of improved RTLS engine and IoT tag using machine learning technique reflecting real environments) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea).

• First Author : (ORCID:0000-0003-2239-6360)Div. of Electronics & Electrical Eng, Dongguk University, weiyiqiao@dongguk.edu, 학생회원
° Corresponding Author : (ORCID:0000-0001-7629-7865)Div. of Electronics & Electrical Eng, Dongguk University, shwang@dongguk.edu, 종신회원
* JMP System, smlee@jmpsystem.com

논문번호 : 201809-282-D-RN, Received September 16, 2018; Revised October 22, 2018; Accepted October 24, 2018

demanding to meet the requirements of 5G services [1]. According to the used signal, these technologies can be mainly divided into two categories[2], one based on radio signals and the other based on none-radio signals. Wi-Fi signal based indoor positioning is the most popular one due to its wide deployment and low cost of Wi-Fi networks. Wi-Fi based positioning technology can be classified into two types, time-space attributes methods and received signal strength (RSS) based methods. The latter is also known as fingerprint method, which is one of the most popular indoor positioning technologies[3]. In Wi-Fi based fingerprint indoor positioning the user location is determined by matching and comparing the real-time RSS measurements with the pre-collected RSS fingerprints. Many machine learning classifiers have been introduced into indoor localization to handle the matching task[4], which equals to a complicated multi-classes classification problem.

Random Forest is one of the most powerful Machine Learning algorithms available today[5]. It can handle both classification and regression task and has been widely used in various applications such as Internet traffic interception, voice and image classification. However, it has been rarely utilized in indoor positioning. As far as we know, currently[6] is the only research which utilized random forest classifier in fingerprint indoor positioning to estimate the user location. In [6], using RSS fingerprinting method, the random forest model of Waikato Environment for Knowledge Analysis (Weka) is utilized to classify 50 target positions with RSSs of 86 deployed APs. The simulation results indicated that the random forest classifier presents positioning accuracy higher than 91%. However, error margin of presented positioning accuracy i.e. positioning precision is not defined. All the results presented in [6] are achieved by simulations so the realistic indoor positioning performance is lack of real-time experimental verification.

In this work, a Wi-Fi based RSS fingerprint indoor positioning system is proposed and implemented. Firstly, RSS of Wi-Fi access points (APs) available at 74 target locations are collected to construct fingerprint dataset for classifier training and test simulation. Then Random Forest classifier has been used to determine the user position. As contrast, other two kinds of popular machine learning classifiers including K-Nearest Neighbors (KNN) classifier, Support Vector Machine (SVM) classifier are also utilized respectively. Modules in Scikit-learn[7] are used to implement all three classifiers in this work. Then the performance of proposed RSS fingerprint indoor positioning system is further verified through real-time indoor positioning experiment.

Ⅱ. Proposed System

In this Section, the overview of proposed system is firstly introduced. The environment and setup are described in Section II.2. The construction of the fingerprint database is introduced in Section II.3. And the most important position prediction program and Random Forest classifier are introduced in Section II.4 and Section II.5 respectively.

2.1 Overview of the proposed system

The proposed system is a server-based[8] and user-active[8] fingerprint indoor positioning. A high-performance positioning server is in charge of the database storage, position prediction (classification task) with our developed algorithms. Fig. 1 describes the actual indoor positioning service delivery by the proposed system. The wireless IoT tag carried by user firstly captures the RSS from different APs in the environment and transfers these measurements with APs’ MAC addresses to the positioning server. After processing received data, the positioning server will inform the user of position decided by the server.

Fig. 2 shows the two phases procedure of proposed Random Forest applied fingerprint indoor positioning system. The offline phase comprises the RSS radio map (fingerprint database) construction and offline Random Forest classifier training. In the online phase, the well trained classifier will make the prediction of the user position according to online RSS measurements transferred by the user

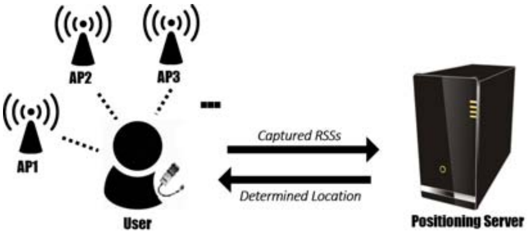

그림 1. 제안된 시스템을 통한 실제 실내 포지셔닝 서비스 제공 다이어그램
Fig. 1. Diagram of the actual indoor positioning service delivery by the proposed system

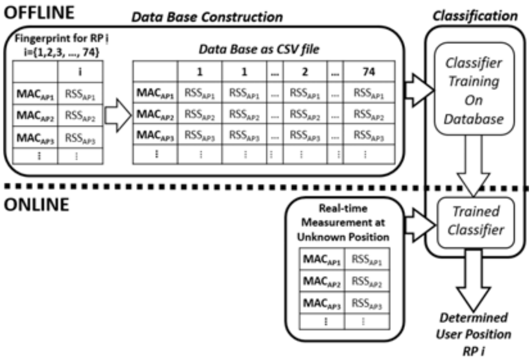

그림 2. 제안된 지문 실내 포지셔닝 시스템 절차
Fig. 2. Procedure of proposed fingerprint indoor positioning system

device in real time. The main task of the fingerprint positioning is to solve complicated multi-classes classification problem where each RP is a single class.

2.2 Environment & Setup

The RSS fingerprint data collection and the final verification experiment are both performed in the 7th floor of the new engineering building at Dongguk University, Seoul, Korea. As shown in Fig. 3, the 52 meters *32 meters target area is divided into target 74 RPs with the interval of 2 meters. About the hardware used in this work, the positioning server is Dell Allienware Model. P31e. The hardware parameters of IoT device used for RSS measurement is summarized in Table 1.

On the software side, the fingerprint database construction program, classification (position prediction) program and online experiment program are all developed in Python language.

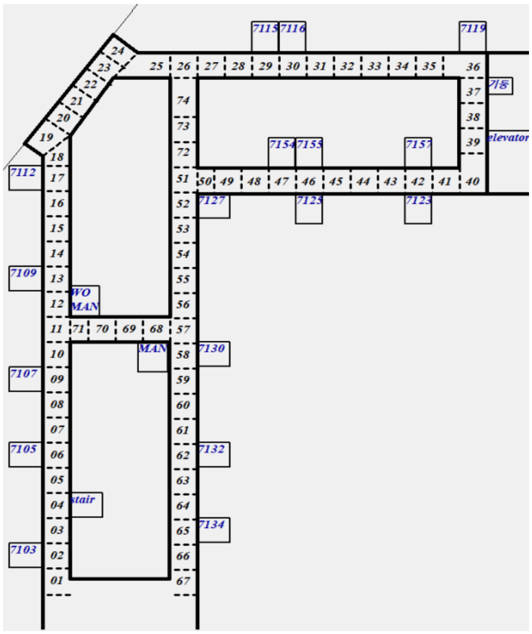

그림 3. 환경 바닥 지도 및 정의된 74개의 참조점
Fig. 3. Environment floor map and defined 74 reference points

표 1. RSS 지문 데이터 수집에 사용되는 IoT 장치 매개변수
Table 1. Parameters of IoT device used for RSS fingerprint data collection

Parameter	Value
Frequency	2.412~2.4835GHz
Wireless standard	802.11bgn
Receiving Sensitivity	-93dBm
Power	5V
TX Power	under 1W
RX Power	under 0.5W
Antenna Gain	1dBi

2.3 Database Construction

As shown in the left side of Fig. 2, firstly, the wireless IoT tag measures the RSS from available APs at each RP. The measurements are combined with APs' MAC addresses and the RP label, then transmitted to positioning server as one fingerprint data example. Considering multipath effect existing in indoor environment, for each RP, 5 fingerprints are collected in forward, back, left and right four directions respectively. After collecting fingerprint data examples at all RPs, totally 10341 fingerprint data are collected. All these examples are saved as one CSV file in the positioning server. Then the dataset will be used for training the machine learning classifier in the positioning server.

2.4 Position Prediction Program

To predict the user position, the main task of positioning server is to solve complicated multi-classes classification problem where each RP is one class. Utilizing Scikit-learn classifier module, a program is developed for positioning server to handle the classification task. Fig. 4 shows the flow chart of the developed program. After loading the aforementioned fingerprint CSV file, the whole fingerprint dataset is split into two subsets, training dataset (5910 samples) and test dataset (4431 samples). Training dataset is used to train the machine learning classifier in offline phase. Test dataset is used for offline test simulation to examine the performance of different classifiers and find the most powerful one. Specifically, RP labels are separated from test data and unlabeled data is input into the trained classifier model. The trained model predicts the RP number when unseen examples are given. Finally, the predictions are compared with separated RP labels to evaluate the prediction precision. Moreover, the trained classifier model can be saved for further experimental verification.

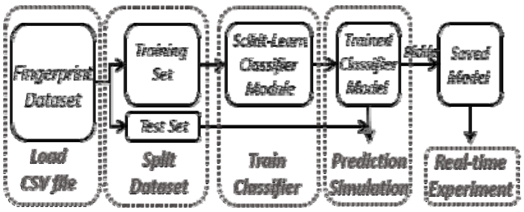

그림 4. 포지셔닝 서버 분류를 위해 설계된 프로그램의 플로 차트
Fig. 4. Flowchart of the designed program for classification in positioning server

2.5 Random Forest Classifier

Random Forest is an ensemble of Decision Trees trained via the bagging method[5]. Decision Tree is a machine learning algorithm with tree-like structure where each node represents a feature, each branch represents a decision rule and each leaf represents an class label. In training phase, to create the Decision Tree, the training dataset is firstly split into two purest subsets by minimizing the following Classification And Regression Tree (CART) cost function[5]:

$$J(k,tk)=Gleft(mleft/m)+Gright(mright/m) \qquad (1)$$

where k is a single feature, tk is its threshold, m is the number of all data examples, mleft/right is the number of examples in left/right subset and Gleft/right measures the impurity of left/right subset. Then the subsets will be further splitted with the same logic. The prediction is made by checking the value of feature while traversing the nodes in the created Decision Tree. Random Forest consists of a group of Decision Trees. All Decision Trees will make decision and the final prediction of Random Forest classifier is the prediction with the most votes[5].

To implement Random Forest classifier, also SVM classifier and KNN classifier, the sklearn.enesmble.RandomForestClassifie, sklearn.svm.svc and sklearn.neighbors.KNeighborsClassifier provided by Scikit-Learn[7] is used respectively. Scikit-learn is a free software machine learning library for the Python programming language and is famous as a simple and efficient tool for data mining and data analysis.

Ⅲ. Simulation & Experiment Results

3.1 Simulation Results

The positioning performance is measured in Accuracy and Success. Accuracy is the percentage of correct RP predictions without margin of error when training data samples are reused as input to the trained classifier model. Success is the percentage of correct position prediction with a margin of error of 2 RPs when test dataset is input. The interval between RPs is 2 meters so the precision of Success is 4 meters.

As mentioned earlier, Random forest consists of a group of Decision Trees. So the number of trees can greatly affect the performance of Random Forest classifier. To investigate this effect and find the appropriate number of decision trees, we tried different number of generated trees, ranging from 10 (default number) to 120. The results are presented in Fig. 5. It can be seen that, when the number of generated trees increases from 10 to 50, the positioning Success also increases significantly but the improvement becomes slower for higher numbers. This is due to that the final prediction of Random Forest classifier is the prediction with the most votes and increasing number of decision trees can make the votes more accurate. When the number of trees beyonds 100, the Success maintains around 94%. Considering it is meaningless to further exchange cost for marginal improvement, the number of generated decision trees is set to 100.

About KNN classifier, classification is computed from a simple majority vote of the k nearest neighbors. The nearest neighbors are decided by calculating distance metrics. So it is important to choose the appropriate distance metrics and find optimal k for fingerprint dataset. parameter tuning is performed with our dataset to achieve the best performance by GridSearchCV module where parameter values for each classifier model are optimized by cross-validated grid-search over a given parameter grid. In the given parameter gird, value set {1,2,3, ..., 50} is for k, and four kinds of distance metrics, Euclidean distance, Chebyshev distance Minkowski distance and Manhattan distance are included. Through GridSearchCV in Scikit-learn, it is found that using Manhattan distance with k equals to 41 gives the highest positioning Success of 82.28% and Accuracy of 40.89%.

About SVM classifier, parameter C for margin violation control and parameter γ in kernel functions give the greatest effect on the performance of the classifier. Using same GridSearchCV method, the optimal C and γ are found to be 100 and 0.01 given a parameter grid, in which {10e-4,10e-3,10e-2,10e-1,1,10e1,10e2,10e3,10e4} is the value set for both C and γ. With this parameter setting, SVM classifier achieved the Accuracy of 98.05% and Success of 85.57%.

The results of performance measurements of Random Forest classifier and the other two classifiers are summarized in Table 2. As the results show, KNN classifier shows the worst performance given both training and test set. Random Forest classifier has the highest test Success. Although SVM classifier shows the best training Accuracy but the test Success is around 8.5% lower than the Random Forest classifier. It means the generalization ability of Random Forest classifier is better than SVM. Random Forest is more capable of handle the instability and variability of RSS of Wi-Fi signals in the complex indoor environment and hence is more powerful for classification task in fingerprint indoor positioning. Because positioning precision is not defined in [6], it is difficult to compare the performance between proposed system and system in [6].

3.2 Experiment Results

Since Random Forest shows the best positioning performance in simulation, the well trained model is saved and used for the online real-time experiment to verify the actual performance of the proposed system. In experiment, IoT tag carried by user captures the real-time RSSs of available AP and send it with MAC addresses to the positioning server. The trained classifier model is load and used to predict user position with received data. In our online experiment, one user carries the IoT tag moved from RP 1 to RP 74. At each RP, RSS measurement and position prediction are both performed 5 times continuously. The predicted positions results are recorded and summarized in the following Table 3 together with the simulation results. The Success defined for positioning simulation is also used as performance measurements in the experiment. The results show that experiment Success is 91% but 3% lower than the simulation result. One reason for this difference is the changing of environment can change the propagation path of Wi-Fi signals and hence greatly affects the RSSs of APs at RPs. The experiment is performed months after fingerprint data collection while test dataset used for simulation is split from

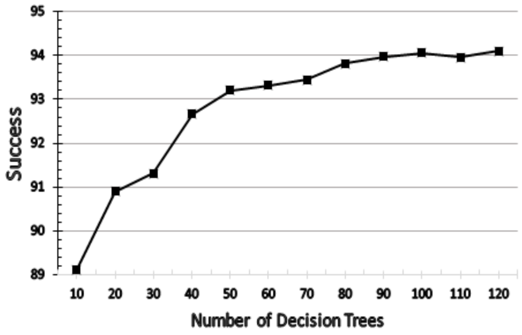

그림 5. 결정 트리의 수량이 다른 랜덤 포레스트 분류기의 위치예측 성능
Fig. 5. Position Prediction performance of Random Forest classifier with different number decision trees

표 2. 3종 기계학습 분류기간의 시뮬레이션 성능 비교
Table 2. Simulation performance comparison between three kinds of machine learning classifiers

Classifier	Accuracy	Success
Random Forest	69.12	94.06
SVM	98.05	85.57
KNN	40.89	82.28

표 3. 랜덤 포레스트 분류기 기반 제안 시스템의 포지셔닝 실험 결과
Table 3. Positioning experiment results of Random Forest classifier based proposed system

	Success
Simulation	94.06
Experiment	91.08

the whole constructed database. The other possible reasons and the way to solve this problem is included in our further research.

Ⅳ. Conclusion

In this work, we proposed a server-based and user-active fingerprint indoor positioning system using Random Forest classifier. We developed algorithms to construct the fingerprint database with IoT device. Then database is used to train machine learning classifier to be able to predict user position. The system is implemented and examined in a real indoor environment with 74 target locations. The simulation results show that Random Forest classifier is more powerful than KNN classifier and SVM classifier with positioning precision up to 94%. The Random Forest classifier applied system is further verified in the online real-time experiment. The proposed system is verified able to provide 4 meters precision indoor positioning with a 91% success rate. Further work includes finding the reason for the gap between simulation and realistic verification experiment, and exploring methods to increase the positioning performance of the proposed system.

References

[1] R. S. Sinha, Y. Wei, and S.-H. Hwang, “A survey on LPWA technology: LoRa and NB-IoT,” *ICT Express*, vol. 3, no. 1, pp. 14-21, 2017.

[2] A. Bensky, *Wireless positioning technologies and applications*, Artech House, 2016.

[3] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, “Indoor fingerprint positioning based on Wi-Fi: an overview,” *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 5, pp. 135, 2017.

[4] Y. Wei and S.-H. Hwang, “A survey on wi-fi based fingerprint indoor positioning technologies,” in *KICS Int. Conf. Commun. 2018*, pp. 627-629, Jeju Island, Korea, Jun. 2018.

[5] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*, O’Reilly Media, Inc., 2017.

[6] E. Jedari, et al., “Wi-Fi based indoor location positioning employing random forest classifier,” *IEEE IPIN 2015*, pp. 1-5, Banff, AB, Canada, Oct. 2015.

[7] F. Pedregosa, et al., “Scikit-learn: Machine learning in Python,” *J. Machine Learning Res.*, vol. 12, pp. 2825-2830, Oct. 2011.

[8] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, “Modern WLAN fingerprinting indoor positioning methods and deployment challenges,” *IEEE Commun. Surveys & Tuts.*, vol. 19, no. 3, pp. 1974-2002, 2017.

위 예 교 (Yiqiao Wei)

2015년 7월 : University of Electronic Science and Technology of China School of photoelectric information 졸업

2016년 9월~현재 : 동국대학교 전자 전기공학과 석사

<관심분야> 무선통신, IoT, 5G

이 상 문 (Sang-Moon Lee)

1994년~2000년 : 삼성전기 선임 연구원

2000년~현재 : (주)제이엠피시스템 연구소장

2015년 : (주)엔티로봇 기술이사

<관심분야> 무선통신, 유선 통신, 머신러닝과 딥러닝 알고리즘 및 구현

황 승 훈 (Seung-Hoon Hwang)

1999년 2월~2005년 : LG전자, 이동통신 연구소 책임연구원

2003년 1월~2005년 2월 : University of Southampton, UK, Visiting research staff

2005년 3월~현재 : 동국대학교 전자전기공학부 교수

2010년 : Stanford University, USA, Visiting professor

2018년 : 한국통신학회 이동통신표준사업 상임이사

2006년~현재 : IEEE Senior member

5G포럼 주파수위원회 / 교통융합위원회 부위원장

<관심분야> 5G 무선통신, IoT, CR, 스펙트럼엔지니어링, 광무선통신, 머신러닝과 딥러닝 알고리즘