

와이파이 핑거프린트 기반 데이터 수집 방법 및 가공 연구

김성현 · 윤창표

경기과학기술대학교

Wi-Fi Fingerprint-based Data Collection Method and Processing Research

Sung-Hyun Kim · Chang-Pyo Yoon

Gyeonggi College of Science and Technology

E-mail : dandyleader@naver.com / cpyoon@gtec.ac.kr

요 약

실내 환경에서 사용자의 위치를 측위하는 다양한 기법들이 있다. 그중 와이파이 핑거프린트 기법은 데이터 수집 단계와 측위 단계로 구분된다. 데이터 수집 단계에서는 해당 위치 주변의 모든 와이파이 신호를 수집하여 리스트 형태로 관리한다. 수집된 데이터가 많을수록 실내측위 정확도가 향상된다. 기존 고품질 데이터 수집 및 관리 방법은 많은 시간과 비용이 소모되고, 기계학습에 필요한 데이터를 추출해 생성할 때 많은 연산이 필요하다. 따라서 한정된 자원 안에서 많은 데이터를 수집 및 관리할 수 있는 방법을 연구한다. 본 논문은 효율적인 데이터 수집 기법과 기계학습에 필요한 학습 데이터 관리 및 생성 기법을 제안한다.

ABSTRACT

There are many techniques for locating users in an indoor spot. Among them, WiFi fingerprinting technique which is widely used is phased into a data collection step and a positioning step. In the data collection step, all surrounding Wi-Fi signals are collected and managed as a list. The more data collected, the better the accuracy of the indoor position based on Wi-Fi fingerprint. Existing high-quality data collection and management methods are time consuming and costly, and many operations are required to extract and generate data necessary for machine learning. Therefore, we research how to collect and manage large amount of data in limited resources. This paper presents efficient data collection methods and data generation for learning.

키워드

Wi-Fi Fingerprint, Indoor positioning, Data collection, Training data generation

1. 서 론

최근 실내 위치에 따라 제공되는 서비스 수요 증가로 인해 다양한 기법을 이용한 사용자 실내 위치 측위 기법이 사용되고 있다. 실내 측위를 위한 여러 가지 알고리즘 중 전파 지문 방식이 가장 보편적으로 사용된다 [1]. 전파 지문 측위 방법은 수집단계에서 측위 데이터를 기반으로 전파 지문을 구축한 뒤 측위단계에서 스마트폰이 수집한 데이터와 비교하여 위치를 판별하는 방식이다.

최근에는 수집된 데이터를 기반으로 인공지능망을 접목한 측위 기법이 연구되고 있다 [2]. 인공지능망 기반의 전파 지문 측위는 적은 양의 학습 데이터를 사용하면 과소적합(Underfitting)을 하기 때문에 고밀도 데이터를 수집하기 위해 많은 자원을 소비해야 된다.

본 논문은 한정된 자원 안에서 많은 데이터 수집을 위해 다중 스마트폰을 활용한 인공지능망 학습 데이터 관리 시스템 구현 방법을 제안한다. 학습 데이터는 측위 환경과 동일한 수집 환경으

로 무선 와이파이 신호 데이터를 수집하기 위해 스마트폰 애플리케이션을 활용한 데이터 수집이 이뤄진다 [3]. 데이터 수집 단계에서는 데이터 관리 서버가 스마트폰으로 수집한 데이터를 데이터베이스에 순차적으로 저장한다. 데이터베이스에 저장된 데이터는 데이터 생성 단계에서 사용된다. 데이터 생성 단계는 와이파이 전파지문 기반의 인공지능망을 학습하기 위한 데이터를 가공하는 단계이다. 이 단계에서 생성된 학습 데이터는 데이터 관리 서버에 요청한 데이터 개수·기간에 의해 다양한 형태로 제공된다.

본 논문의 전체적인 구성은 다음과 같다. II장

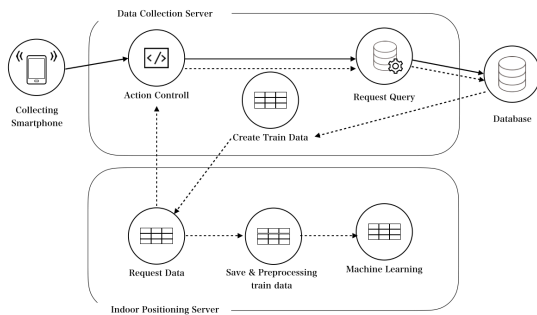


그림 1. 제안 시스템 구조
Fig. 1. Suggested System Structure

에서는 제안 시스템 구조를 기점으로 파트별 시스템 구현을 단계적으로 설명한다. III장에서는 결론 및 향후 연구를 제시한다.

II. 데이터 관리 시스템

2.1 제안 시스템 구조

본 논문은 수집 데이터 관리에 대한 시스템 구조를 그림 1과 같이 구현하였다. 제안한 시스템 구조는 3가지 형태로 구분한다. 신호 탐색 기능을 탑재한 스마트폰 애플리케이션으로 데이터 수집 부분을 구현한다. 애플리케이션은 Media Access Controller Address (MAC Address)에 해당되는 Received Signal Strength Indicator (RSSI)를 페어 (pair) 형태로 데이터를 수집한다. 수집된 데이터와 위치정보를 데이터베이스에 저장될 수 있도록 서버에 전달한다.

수집 데이터 관리 서버는 다중 스마트폰에서 수집한 데이터를 관리하고 인공지능망의 학습 데이터를 생성한다. 서버는 스마트폰 애플리케이션이 송신한 데이터의 위치정보에 따라 건물 별 전파 지문 데이터베이스를 구현한다. 구현된 전파 지문 데이터베이스를 토대로 인공지능망 학습 서버의 개수·기간 요청에 따라 다양한 형태의 학습 데이터가 생성된다. 기계 학습 데이터 생성이 완

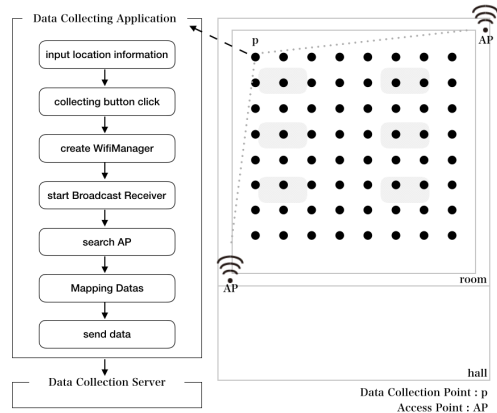


그림 2. 데이터 수집 애플리케이션
Fig. 2. Data Collecting Application

료되면 학습 서버에게 전달한다. 인공지능망 학습 서버는 데이터 관리 서버로부터 수신된 데이터로 학습을 진행한다.

2.2 안드로이드 스마트폰을 활용한 데이터 수집

와이파이 신호 데이터 수집을 위한 애플리케이션은 그림 2와 같은 과정을 통해 진행된다. 측위 지점에서 측정 장소에 대한 위치 정보를 입력한 후 수집 버튼을 클릭하여 데이터를 수집한다. 측정 장소에 대한 위치 정보 입력은 수집데이터 관리의 용이성과 기계학습 데이터 생성 시 필요한 데이터만 선별하여 제공하기 위해 구분자로서 한다. 위치정보 입력은 건물 내부이기 때문에 GPS를 이용해 디바이스 위치를 확인할 수 없어 직접 입력하는 것을 권장한다. 매회 수집된 데이터들은 다수의 Access Point (AP)가 탐색되기 때문에 데이터들을 Key-Value 형태로 MAC Address와 RSSI를 한 쌍으로 매핑한다. 변환된 데이터들은 HyperText Transfer Protocol (HTTP)를 통해 약속된 액션 값을 포함하여 데이터 관리 서버에 전송한다. 서버는 전송받은 데이터의 액션 값을 구분하여 작업을 처리한다.

위 과정을 반복하여 그림 2와 같이 격자 형태의 지점에서 학습 데이터를 수집한다. 인공지능망 기반 와이파이 전파 지문 기법 연구 시 지점간의 거리가 촘촘할수록 공간에 대한 많은 데이터를 학습 데이터로 사용할 수 있기 때문에 향상된 측위 결과를 도출해 낼 수 있다. 표본 데이터 수집 이후 주기적으로 데이터 수집을 진행하고 측위에 혼동이 발생할 여지가 있는 벽과 문 앞 등의 지점에서 추가적인 데이터를 수집하여 공간 특성 변화에 대응한다.

2.3 데이터 관리 서버의 수집 데이터 저장

제안된 수집 데이터 관리 서버의 설계는 그림 3과 같다. 관리 서버는 수집 데이터 저장 부분과

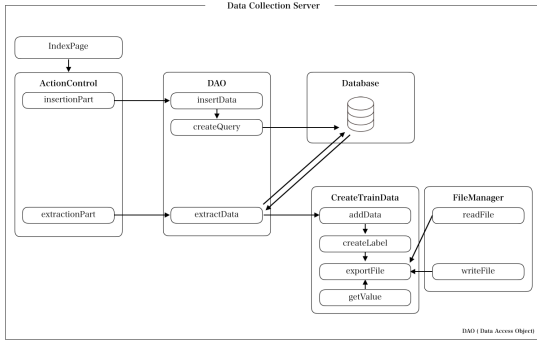


그림 3. 데이터 수집 서버 설계도
Fig. 3. Data Collection Server Structure

기계 학습 데이터 생성 부분으로 나뉜다. 서버는 요청 따른 작업을 구분하기 위해 액션 값을 사용한다. 데이터베이스는 관리 서버로부터 요청받은 데이터를 축적한다. 데이터베이스는 수집 위치 정보를 관리하는 테이블과 수집 위치에 따른 와이파이 신호 데이터들을 관리하는 테이블로 구성된다. 수집 위치 정보는 수집데이터 관리 테이블에서 왜라키 (Primary Key)로 사용된다. 왜라키를 사용하면 기계 학습 데이터를 생성할 때 원하는 인공지능 기반 학습 서버의 요청 사항을 빠르게 검색하여 제작할 수 있다.

수집 데이터 삽입과 학습데이터 생성 요청은 서버를 통해 데이터베이스에 접근이 이루어진다. 이때 동적 쿼리 (Dynamic Query) 삽입을 통한 데이터베이스 공격이 발생할 가능성이 있다. 동적 쿼리 삽입에 대한 취약점을 막기 위해서 행정 안전부가 발간한 Java 시큐어 코딩 가이드를 사용하였다. Java 시큐어 코딩 가이드를 기반으로 SQL 삽입 (SQL Injection)에 대비한 보안 코드 방식을 적용하여 쿼리를 생성한다.

2.4 데이터 관리 서버의 학습 데이터 생성 부분

기계 학습 데이터 생성은 데이터 추출과 파일 형식 지정, MAC Address 컬럼 설정, MAC

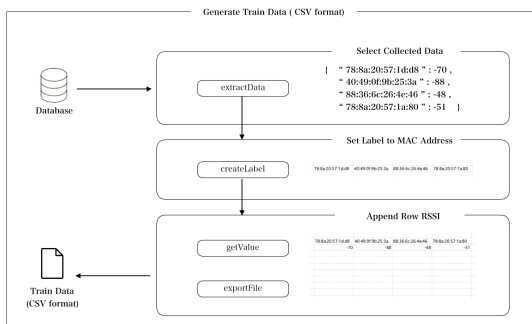


그림 4 기계학습 데이터 생성 과정
Fig. 4. Generate Process of Learning Data

Address에 대한 RSSI 삽입으로 그림 4와 같은 과정으로 구현한다. 파일형식은 Comma-Separated Values 형식을 기반을 사용한다. 인공지능 기반 학습 서버는 학습 데이터 생성을 위해 지정된 양식(Form)을 작성하여 데이터 관리 서버로 요청한다. 데이터 관리 서버는 요청사항을 토대로 데이터베이스에서 수집 기간과 건물, 개수를 지정하여 데이터를 추출한다. 추출된 데이터는 Key-Value 형식으로 매핑되어 있는 것을 확인할 수 있다. 매핑된 데이터들은 Iterator 패턴을 이용하여 MAC Address 추출 후 컬럼을 제작한다. 제작된 컬럼을 기준으로 RSSI를 레코드로서 삽입한다. 생성된 학습 데이터는 인공지능 기반 학습 서버로 전송되어 학습이 진행된다.

III. 결 론

본 논문은 다수의 스마트폰으로 수집한 와이파이 신호 데이터들을 관리하고 인공지능 기반의 학습데이터 조건에 맞춰 데이터를 구성할 수 있는 데이터 관리 서버 구현을 제안한다. 데이터 수집 단계에서 다수의 스마트폰으로 서로 다른 위치에 대한 RSSI를 동시에 수집할 수 있어 수집에 소요되는 시간을 단축할 수 있다. 데이터 생성 단계에서 데이터 관리 서버가 수집 환경에서 감지되는 AP의 변동을 반영하기 때문에 사용하지 않는 AP를 제거하고 새로 감지된 AP를 추가 할 수 있다.

와이파이 지문 기법 기반의 실내측위 기술은 AP의 특성상 지속적인 데이터 수집으로 전파 지문을 업데이트해야 한다. 수집 단계에서 수집된 데이터 외에 측위 단계에서 수집한 데이터를 가공해 지속적인 학습데이터 생성에 대해 추가 연구가 진행되어야 한다.

References

- [1] M. N. Husen and S. Lee, "Indoor human localization with orientation using WiFi fingerprinting", in *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication (ACM)*, NewYork: NY, p. 109, 2014. Jan.
- [2] C. U. Park, H. G. Shin and Y. H. Choi, "A Parallel Artificial Neural Network Learning Scheme Based on Radio Wave Fingerprint for Indoor Localization.", in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, Pargue, Czech Republic, pp. 794-797, 2018. July.
- [3] N. T. Thuong, H. T. Phong, D. T. Do, P. Van

Hieuw and D. T. Loc, "Android application for WiFi based indoor position: System design and performance analysis." in *2016 International Conference on Information Networking (ICOIN)*. IEEE, Kota Kinabalu, Malaysia, pp. 416-419, 2016. Jan.