

Research Article

Radio Environment Map Construction by Kriging Algorithm Based on Mobile Crowd Sensing

Zhifeng Han , Jianxin Liao , Qi Qi , Haifeng Sun, and Jingyu Wang 

Institute of Network Technology, Beijing University of Posts and Telecommunications, 100088, China

Correspondence should be addressed to Zhifeng Han; zhifeng.han@gmail.com

Received 25 October 2018; Revised 28 December 2018; Accepted 13 January 2019; Published 3 February 2019

Academic Editor: Ilsun You

Copyright © 2019 Zhifeng Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the IoT era, 5G will enable various IoT services such as broadband access everywhere, high user and devices mobility, and connectivity of massive number of devices. Radio environment map (REM) can be applied to improve the utilization of radio resources for the access control of IoT devices by allocating them reasonable wireless spectrum resources. However, the primary problem of constructing REM is how to collect the large scale of data. Mobile crowd sensing (MCS), leveraging the smart devices carried by ordinary people to collect information, is an effective solution for collecting the radio environment information for building the REM. In this paper, we build a REM collecting prototype system based on MCS to collect the data required by the radio environment information. However, limited by the budget of the platform, it is hard to recruit enough participants to join the sensing task to collect the radio environment information. This will make the radio environment information of the sensing area incomplete, which cannot describe the radio information accuracy. Considering that the Kriging algorithm has been widely used in geostatistics principle for spatial interpolation for Kriging giving the best unbiased estimate with minimized variance, we utilize the Kriging interpolation algorithm to infer complete radio environment information from collected sample radio environment information data. The interpolation performance is analyzed based on the collected sample radio environment information data. We demonstrate experiments to analyze the Kriging interpolation algorithm interpolation results and error and compared them with the nearest neighbor (NN) and the inverse distance weighting (IDW) interpolation algorithms. Experiment results show that the Kriging algorithm can be applied to infer radio environment information data based on the collected sample data and the Kriging interpolation has the least interpolation error.

1. Introduction

In the IoT networks, 5G technology is characterized by higher bit rates with more than 10 Gigabits per second as well as by more capacity and very low latency, and it will leverage novel technological concepts to meet the “anywhere and anytime” requirements of IoT devices. With the rapid development of IoT devices amount, the demand for wireless spectrum resources is increasing. In order to dynamically plan the spectrum resources to improve the utilization of radio resources to provide well control of IoT devices’ access control, we can build radio environment map (REM) to collect and understand the radio information. REM can offer multidomain environmental information, such as geographical features, available services, spectral regulations, locations and activities of radios, relevant policies, and experiences [1].

However, the primary problem of constructing REM is how to collect large scale of data. Currently, most of the REMs are aimed at small scale and applied to specific applications. And the universal methods to build a REM are by deploying sensors in a certain environment to collect the sensing data. However, the REM is applied to dozens of different kinds of networks and applications, which makes the networks and applications have to collect data separately [2, 3]. Besides, the same data can hardly be shared and reused among different applications, resulting in duplication of data collection and a waste of resources. Therefore, it is of great significance to construct a large scale and universal REM, which can integrate data sources of radio environment and avoid the cost of the reconstructing database [4]. Mobile crowd sensing (MCS) is an effective solution to solve this problem, which is a novel emerging paradigm that leverages the smart devices

carried by ordinary people to collect information and has facilitated many sensing applications, such as environment monitoring, traffic detection, social interaction, and public information sharing [5]. MCS can be applied to collect radio environment information in the sensing area. In order to characterize environmental information comprehensively, recruiting adequate ordinary users with smart devices to participate in radio environment information collection is needed. Compared with the traditional data collection technologies, MCS collects the environment information by built-in sensing modules in the mobile terminals, and it has the properties of mobility, the ubiquity of nodes, the powerful storing, and computing ability [6, 7].

Wireless network signals are all electromagnetic waves, whose transmission and attenuation are complex process. Therefore, in this paper, we only analyze the transmission and attenuation processes of electromagnetic waves in space entropy in ideal conditions. Under ideal conditions, the propagation process is free from any obstruction and without any multipath propagation. Then the propagation model of space electromagnetic waves is a free space propagation model. According to the pattern of wireless electromagnetic wave transmission in free space, spatial interpolation algorithm can be applied to restore the uncollected radio environment information data of the sensing area. The Kriging interpolation algorithm has been widely used in geostatistics principle for spatial interpolation but is not broadly used in wireless network area. Kriging spatial interpolation algorithm estimates unknown point data and not only considers the relative positions of estimated points and known sample points, but also considers the relative positional relationship between all sample points. In this paper, we proposed to use Kriging interpolation algorithm to infer the uncollected radio environment information based on the collected sample data.

In this paper, we proposed to apply the MCS to collect the radio information. Furthermore, to address the problem of the incomplete radio environment information caused by the inadequate sensing data, we proposed to apply the Kriging interpolation algorithm to infer the uncollected radio environment information with the collected sensing data. Our contributions are as follows:

- (i) We propose a REM prototype system based on MCS, where the ubiquitous, massive, and high dimension REM-related data can be sensed and collected by the terminals carried by mobile users.
- (ii) We propose to apply Kriging interpolation algorithm to infer the uncollected radio environment information caused by the target area being not covered by the participants.
- (iii) We set up experiments to collect the sample of the radio environment information and infer the missing radio information of the target area. The results show that the Kriging interpolation algorithm can infer the missing radio information and has the least interpolation error.

The rest of the paper is organized as follows. In Section 2, the related works are introduced. Section 3 outlines the architecture of the REM based on MCS. In Section 4, the Kriging interpolation algorithm is introduced. The simulation results are illustrated in Section 5. Section 6 presents the conclusion.

2. Related Works

Building REM needs a large number of sensors and kinds of radio environment information, which is a great challenge. At present, data collection methods for REM can mainly be categorized into three types. First is integrating or accessing the related information directly from existing databases, estimating radio propagation characteristics by software tools, and leveraging cognitive radios devices or networks to sense data. Gathering data from the existing database is a relatively convenient way, while the data updating time depends on the updating period of the underlying database. Moreover, the historical information is not stored in the underlying database. Riihijärvi et al. take vantage external datasets to build REM, but the update cycle of the external datasets is very long which makes datasets unable to meet the real-time requirement of REM [8]. Constructing REM in this way makes it difficult to satisfy the upper-layer applications with the requirement for real-time and historical information. Second, the way to characterize and estimate the properties of radio transmission based on software is to calculate the signal attenuation by modeling so that we can better plan the radio environment [9, 10]. The model in [11] clearly gives a solution to the signal diffraction problem caused by the occlusion, but this requires an accurate vector model of all three-dimensional structures, with limited data and resolution in most experimental environments. It cannot be applied to applications that require high accuracy. The above-mentioned estimation method usually provides limited data, bad accuracy of the data. Third, the method based on wireless device or external network mainly uses the information sensing ability of heterogeneous spectrum sensor network to collect data [12, 13]. In terms of data collection, MCS refers to the sensing paradigm in which mobile users with sensing and computing devices are tasked to collect and contribute data in order to enable various applications [14]. It combines people-centric sensing and crowdsourcing so that a great number of ordinary users with smart devices can cooperate with each other to form a sensing network and deliver the sensing tasks [5]. Then participants can upload the sensing data to the MCS platform. The development of MCS has resulted in various novel sensing applications. Some typical examples include the air quality inspection application Common Sense for air quality monitoring [5] by the University of California Berkeley, the Creek Watch application to evaluate city water resources [15] by IBM, and the Nericell system [16] by Microsoft to monitor road and traffic condition implemented by piggybacking on smartphones that users carry with them in normal course. MCS has attracted much attention from researchers due to advantages such as ubiquitous sensor nodes, good participant mobility, low maintenance cost, and rich sensing data types. Hence, MCS can be applied to collect the radio environment information.

MCS can be applied to collecting large scale data due to the properties of mobility, the ubiquity of sensing nodes. However, limited by the budget, there are not enough participants recruited to join the data collection, which makes the radio environment information incomplete. To build the REM, complete radio information is needed. A lot of works have been done about inferring the missing data according to the sample data in many research fields. Talvitie et al. investigate spatial interpolation and extrapolation algorithms for construction of fingerprint databases [17]. Lacking knowledge about the beacon locations, measurement at an unknown point is interpolated based on actual measurements in the surrounding. There are several interpolation algorithms considered in [17], which include linear interpolation based on Delaunay triangulation, the nearest neighbor (NN), and the inverse distance weighting (IDW) to name a few. The results show that location accuracy is enhanced by utilizing constructed databases comparing to the incomplete database. Grimoud et al. use an iterative process to obtain the REM based on Kriging interpolation to reduce the measurement data required [18]. Umbert et al. apply Kriging and a modified version of the Inverse Distance Weighted (IDW) algorithm to build a REM of an outdoor TV spectrum resources [19]. Hence, there is a spatiotemporal correlation between radio environment data, and the Kriging interpolation algorithm can be applied to infer the missing radio environment information data. Here, we use Kriging interpolation algorithm to infer the missing radio environment information data according to the collected sample data.

3. The REM Based on MCS Architecture

In this section we will introduce the REM based on MCS. First, we present the system architecture and discuss the system components of the radio environment information data collection platform. Second, we introduce the data collection process used to collect the radio environment information data.

3.1. Radio Environment Information Collection System Architecture

3.1.1. Radio Environment Information Collection Platform. Figure 1 shows the overview of our system based on MCS. As shown in Figure 1, from bottom to upper layer, the system includes data sensing layer, data collection layer, data processing layer, data analysis layer and visualization layer. In the data sensing layer, a large number of mobile terminals constitute the mobile crowd sensing network, and they play the role of data sensing by running our data collecting APP named wireless detect. The mobile terminals upload the sensing data to our cloud servers via Wi-Fi/3G/4G networks. The data collection layer is mainly responsible for receiving data, node selection, task allocation, and making incentive mechanism to recruit enough interested nodes to participate in the sensing tasks. The data preprocessing includes arranging the data format and data fusion. The data analysis layer is responsible for the statistical analysis and calculation of the radio environment relevant parameters. At

last, the visualization layer shows the REM relating results in the forms of the field strength map, heat map, and some other maps.

Our proposed architecture involves various functional blocks, communicating via well-specified interfaces. To establish a complete radio environment map, the fundamental problem is the collection of a large number of data with complex types and data processing and visualization. Our system consists of five different function modules: data sensing, data collection, data processing, data analysis, and visualization; each of them has its own function.

Data sensing module is operated by the MCS network, which is organized by mobile terminals carried by mobile users. When a mobile user receives a data sensing task, it will determine whether the user is involved in the task. If so, it will collect the required data by the sensing module embedded in the terminal. Moreover, it also uploads data to the web server by different types of network accessing technologies like Wi-Fi/3G/4G. Our system includes perception of user-uploaded data and calls, mobile phone map API, real-time construction of heat map, and signal strength map. Users can use wireless detection real-time view of the environment in which the radio spectrum resources are used.

Data collection mainly includes area partition, incentive mechanism, nodes selection, task distribution, data storage, and data distribution. The area partition is designed to identify whether a sensing task refers to a geographical location or is based on some social relationships. In our system, we divided it into regional division and business division. The incentive mechanism is used to reduce the cost of the platform as well as attracting enough sensing users. Furthermore, node selection mechanism needs to select the appropriate node for the data sensing and also needs to assign the sensing nodes to the corresponding sensing tasks if there is more than one task.

Data processing module mainly includes two modules: data preprocessing (filtering and cleaning) and data fusion, which is implemented by the MapReduce workflow. The data processing flow is as follows. Firstly, the Avro in the data fusion module compresses various types of formats of the data and merges massive small files into large files to improve the efficiency of MapReduce. Secondly, as the raw data is varying in data types, the data cleaning and filtering can play an important role to remove the noise and interference such as error data. Thirdly, these data are processed by sever cluster, and the processing results are stored in data center.

Data analysis is responsible for the statistical analysis and calculation after the data preprocessing. In order to exhibit the radio environment on the map, it needs to perform analysis and calculation to get the related parameters such as the channel occupation, frequency band occupancy, and background noise intensity.

The visualization module is responsible for the REM-related data parameters exhibition. We designed the visualization for the REM properties. The system can show the Wi-Fi signal coverage, cellular signal coverage heat map, and Wi-Fi channel occupation ratio map. The visual REM

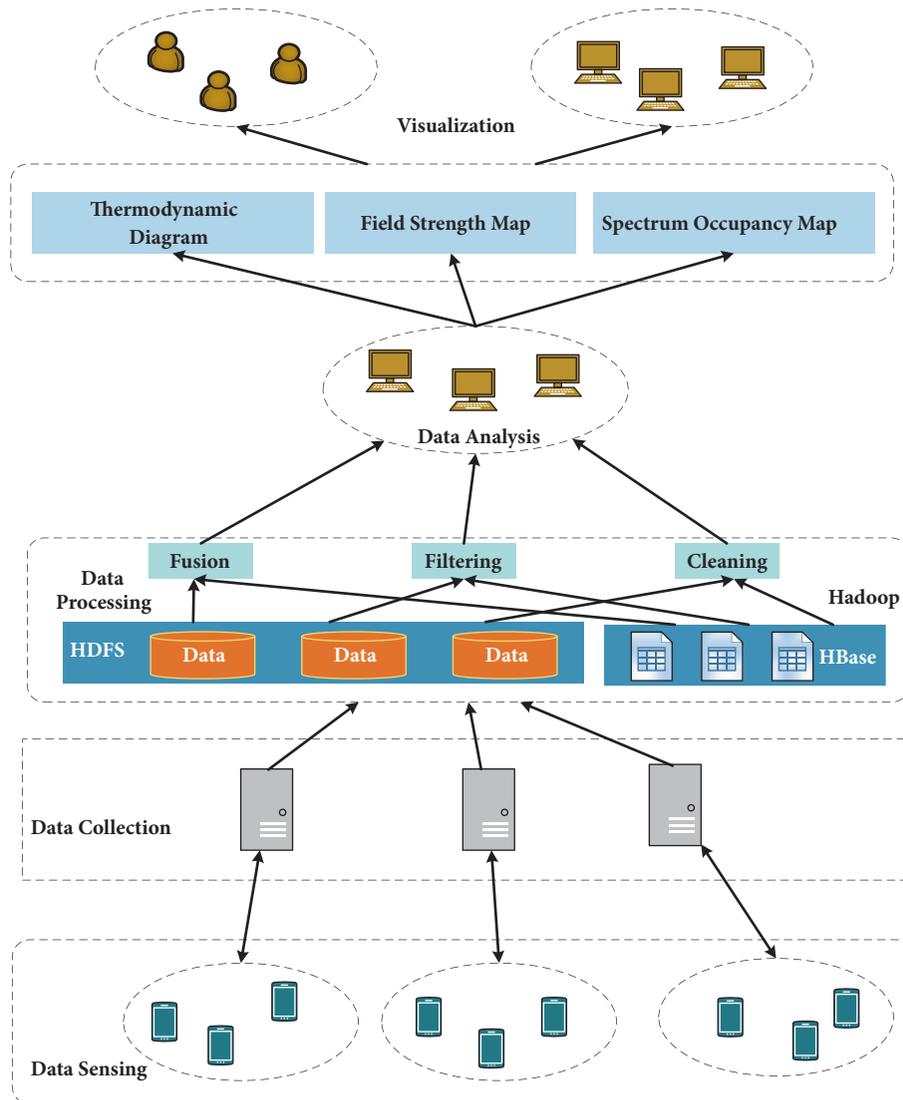


FIGURE 1: Radio information collection architecture based on MCS.

makes it easy to identify the radio environment of the target area.

3.2. Radio Environment Information Collection Process. In [20], the author proposed 4WIH model in mobile sensing and divided the MCS life cycle into four phases, which is shown in Figure 2: task creation, task assignment, individual task execution, and crowd data integration according to the MCS life cycle. Next, we will discuss the following key design issues: REM task creation, REM task assignment, participants recruiting, and participants' selection.

The task creation specifies the sensing timing and coverage area for the REM. In our system, the web server releases the sensing tasks to the users who are interested in the data collection task. REM supports long spatiotemporal information for the upper-layer applications, so the sensing time is continuous.

In our REM task assignment stage, the system is responsible for recruiting and selecting participants for the MCS task. Correspondingly, this stage includes participants recruiting, participants' selection, and incentive mechanism. We choose the well suited participants to join the sensing task to collect the radio environment information, and reward them for the high quality sensing data. The purpose of participants recruiting is to encourage enough people to join the sensing task and get more radio environment data. However, limited by the budget of the platform or the human mobility, only part of the participants can join the radio environment information sensing task. Then the radio environment information data is incomplete; we will talk about the solution later.

In sensing task execution, participants conduct sensing tasks and upload the sensed data to the MCS platform. The participants receive the sensing tasks and then collect the radio environment data. The selected participants are

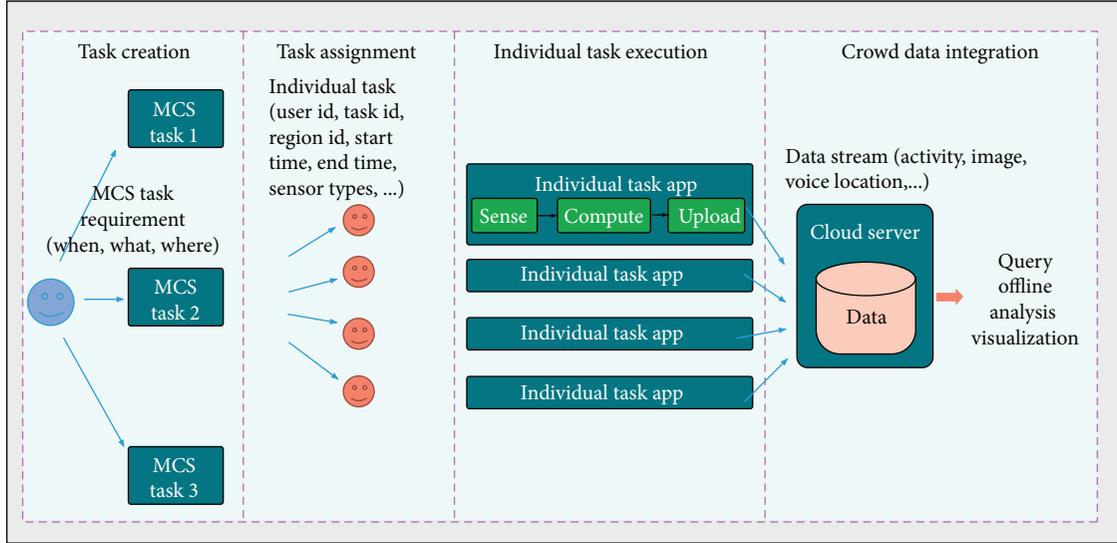


FIGURE 2: The four-stage life cycle of the mobile crowd sensing process [20].

distributed in the target places collecting data. After radio environment data collection the participants upload the data to the MCS platform server by cellular networks (3G/4G) or WLAN.

During the data integration, the main issue is to achieve the MCS task that is to process and analyze the raw data received from mobile terminals and visualize the required results eventually.

4. Missing Data Inference by Kriging

4.1. Related Definition. In this section we will introduce how to infer the missing radio environment information data by Kriging interpolation algorithm. The whole process is shown in Figure 3, which can be divided into two steps. First, we analyze the distribution of sample points distributed in the sensing area and propose a variogram model to reflect the spatial structure characteristics and distribution characteristics of the variables. Second, we use Kriging algorithm to calculate the missing data according to the collected sample radio environment information data. Third, some basic concepts and related definitions are introduced, and all the parameters are listed in Table 1.

Definition 1 (variogram). In spatial statistics the theoretical variogram $2\gamma(s_1, s_2)$ is a function describing the degree of spatial dependence of a spatial random field or stochastic process $Z(s)$. Given an area of interest $G \subset R^n$, the mean of RSS value at a location x_i is considered as a random variable (RV) \bar{Z}_i . Then, the mean of RSS values over the area can be represented by a random field (RF), which is a collection of spatial RVs, $\{\bar{Z}_i | X_i \in G\}$.

Definition 2 (stationary process). Formally, let $\{X_t\}$ be a stochastic process and let $F_X\{x_{t_1+\tau}, \dots, x_{t_k+\tau}\}$ represent the cumulative distribution function of the unconditional (i.e., with no reference to any particular starting value) joint

TABLE 1: Description of parameters in the equations.

Symbols	Descriptions
Z_i	Random variable
$\delta_{i,j}$	Difference between two neighbor points
$E[\delta_{i,j}]$	Mathematical expectation
$\gamma(h)$	Variogram function
$N(h)$	Number of pairs of observations
h	The separation distance
$\gamma_{i,j}$	Variogram value
G	A set representing the area of interest
α, β	Fitting parameters with a strict constraint
z_j^i	RSS value received from the j th beacon

distribution of $\{X_t\}$ at times $t_1 + \tau, \dots, t_k + \tau$. Then, $\{X_t\}$ is said to be strictly (or strongly) stationary if, for all k , for all τ , and for all $t_1, \dots, t_k, F_X(x_{t_1+\tau}, \dots, x_{t_k+\tau}) = F_X(x_{t_1}, \dots, x_{t_k})$.

4.2. Problem Formulation

- (1) Analysis of the distribution of sample points:

Given the sample radio environment information data D_s of the area $G \subset R^n$, we need to use the variogram function to analyze the distribution of sample data in the sensing area. If the sample data only depends on the distance h between the sample data points, we can use Kriging interpolation to infer more data.

- (2) Kriging interpolation process:

Given the sample radio environment information data D_s of the area $G \subset R^n$, we need to infer the complete radio environment information data D_c to build the REM.

4.3. The Variogram. First, some basic concepts and definitions of the Kriging interpolation algorithm are introduced as follows:

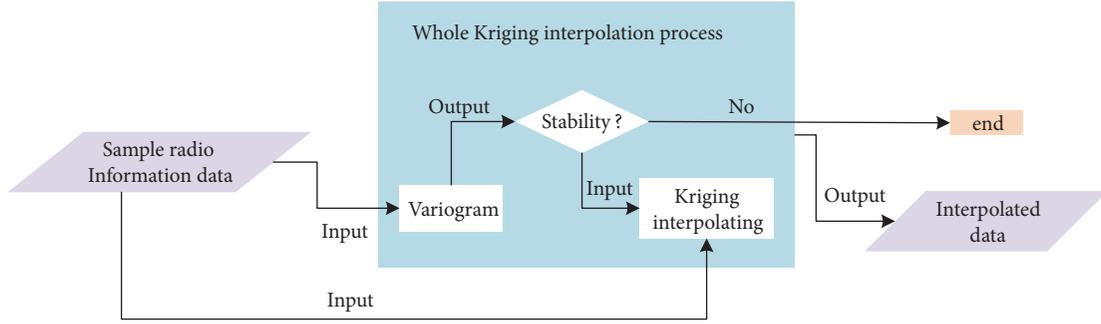


FIGURE 3: The whole process of inferring more radio information data by Kriging interpolation algorithm.

The variogram is often applied in the statistical process of geolocation-related information. variogram can describe the structural change and distribution of variables in the geospatial space. Assume that the value of sample data point x in the sensing area is $Z(x)$ and the sample data value of point $x + h$ is $Z(x + h)$. Then half of the variance of the difference between the values at the two points is defined as the variation of $Z(x)$ at position x . The function can be expressed as

$$\begin{aligned} r(x, h) &= \frac{1}{2} \text{var} \{ [Z(x) - Z(x + h)]^2 \} \\ &= \frac{1}{2} (E[Z(x) - Z(x + h)]^2 \\ &\quad - \{E[Z(x) - Z(x + h)]\}^2) \end{aligned} \quad (1)$$

where $r(x, h)$ is the variogram and $Z(x)$ and $Z(x + h)$ are sample attribute values of the variables at points x and $x + h$ in the target area, respectively. h is the distance between x and $x + h$, and $E\{[Z(x) - Z(x + h)]^2\}$ is the mathematical expectation. In the variogram function, when the increment $Z(x)$ of the variable $[Z(x) - Z(x + h)]$ of the target area satisfies the following two conditions, it is said that $Z(x)$ satisfies the second-order stationarity.

- (i) First, the sample radio environment information data have to satisfy the rules, which are listed as follows:

$$E[Z(x)] = E[Z(x + h)] = m, \quad \forall x, \forall h \quad (2)$$

means the target regionalization variable does not have obvious characteristics in terms of space and fluctuates around m .

- (ii) Second, in the entire target area, the covariance function of $Z(x)$ exists and is stable; namely,

$$\begin{aligned} \text{cov}[Z(x), Z(x + h)] &= E[Z(x)Z(x + h)] \\ &\quad - E[Z(x)]E[Z(x + h)] \\ &= E[Z(x)Z(x + h)] - m^2 \\ &= C(h). \end{aligned} \quad (3)$$

The attribute value $Z(x + h)$ in the target area has no relationship with the position point x and is only related to

h , which means the value is relative to the relative position and does not depend on the absolute position. According to the above two conditions, it can be concluded that the target regionalization variable is strictly second-order stationary in the target area. However, since it is difficult to satisfy a strict second-order stationary state in real life, the condition that satisfies the strict second-order stationary state is weakened to obtain an intrinsic assumption, also called an intrinsic assumption. Similarly, when the increment $[Z(x) - Z(x + h)]$ of the target regionalization variable $Z(x)$ satisfies the following two conditions, it is said to satisfy the intrinsic assumption:

- (i) First, the sample radio environment information data have to satisfy the rules, which are listed as follows.

$$E[Z(x) - Z(x + h)] = 0 \quad (4)$$

- (ii) Second, the variance function of the target regionalized increment $[Z(x) - Z(x + h)]$ exists and is stable; the variance function of the increment does not depend on x , then

$$\begin{aligned} \text{var}[Z(x) - Z(x + h)] \\ &= E[Z(x) - Z(x + h)]^2 - \{E[Z(x) - Z(x + h)]\}^2 \\ &= E[Z(x) - Z(x + h)]^2, \quad \forall x, \forall h. \end{aligned} \quad (5)$$

When the variation of the target area satisfies the weak second-order stationary or intrinsic assumption, due to $E[Z(x) - Z(x + h)] = 0$, the half-difference function can be expressed as follows.

$$(r, h) = \frac{1}{2} E[Z(x) - Z(x + h)]^2 \quad (6)$$

At this time, the increment $[Z(x) - Z(x + h)]$ of $Z(x)$ is only related to the distance between two points. The above variogram function is a theoretical variogram function. In actual operation, multiple sample data needs to be divided into multiple pairs for calculation, like $\{Z(x_i), Z(x_i + h)\} (i = 1, 2, \dots, N(h))$, where $N(h)$ is the number of pairs of points in

the sample data divided by h . The extremes of the variogram can be calculated in the following way.

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (7)$$

In (7), h is the distance between the sample point and the point to be estimated, and $N(h)$ is the number of samples used to calculate the variogram of the sample between $(x_i, x_i + h)$. After the above steps, the analysis of the distribution characteristics of sample points in the target area has been completed. However, in order to estimate the unknown value of the target area variable, the fitting of the convenience function point of the actual sample is also called the theoretical variogram model. The theoretical model of the variogram is to abstract the experimental variogram and then use it to calculate the Kriging interpolation.

The empirical variogram contains values at a limited number of h . To estimate the measurements at unknown locations, access to the value of h between the scattered points in the empirical variogram is required. Hence, a mathematical model is selected to be fitted in the empirical variogram. This model is frequently chosen from spherical model, exponential model, Gaussian model, power model, and linear model. We choose the sample radio environment data and input them to the Matlab. Then we can use the fitting function to find a mathematical expression.

4.4. Kriging Interpolating. Once the variogram is obtained, values at unknown locations can be estimated based on known data points. Mathematically, this problem can be regarded as a spatial interpolation problem. Assuming that the target area to be studied is A , the variable in the target area is $\{Z(x) \in A\}$, where x represents a position in the target area. The sample value of $z(x)$ in the target area $x_i (i = 1, 2, \dots, n)$ is $z(x_i) (i = 1, 2, \dots, n)$. Then the value $z(x_0)$ at the point x_0 to be estimated is the weighted sum of the known n point sampling values:

$$z(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (8)$$

where $\lambda_i (i = 1, 2, \dots, n)$ is the weight coefficient of the known sample point. Due to the fact that $z(x)$ satisfies the second-order stationary assumption when analyzing the distribution of sample points in the target area, then

- (i) there is a mathematical expectation for variable $z(x)$, and the expected value is a constant, $E[z(x)] = m$;
- (ii) there is a covariance function for the variable A ; that is, the value of the point to be estimated in the target area is only related to the distance B between the positions of the known sample points; then,

$$\begin{aligned} \text{cov}\{z(x), z(x+h)\} &= E[z(x) \cdot z(x+h)] - m^2 \\ &= C(h). \end{aligned} \quad (9)$$

According to the unbiased requirements of the interpolation available,

$$E[z^*(x_0)] = E[z(x_0)]. \quad (10)$$

Then,

$$\sum_{i=1}^n \lambda_i = 1. \quad (11)$$

Then under the condition that $z(x)$ is second-order stationary, the calculation process of the estimated variance can be performed by the following method.

$$\begin{aligned} \sigma_E^2 &= E[z^*(x_0) - z(x_0)]^2 - \{E[z^*(x_0) - z(x_0)]\}^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{i,j} - 2 \sum_{i=1}^n \lambda_i C_{i,0} + C_{0,0} \end{aligned} \quad (12)$$

In order to minimize the variance of unbiased estimates, $\min\{\text{var}[z^*(x_0) - z(x_0)] - 2\mu \sum_{i=1}^n (\lambda_i - 1)\}$, and then we can get the equations for the weighting coefficients λ_i in (8).

$$\begin{aligned} \sum_{i=1}^n \lambda_i \text{cov}(x_i, x_j) + \mu &= \text{cov}(x_0, x_i) \\ \sum_{i=1}^n \lambda_i &= 1, \quad i = 1, 2, \dots, n \end{aligned} \quad (13)$$

The equations are the Kriging equations. In addition, (8) is written in matrix form:

$$[K][\lambda] = [M] \quad (14)$$

where

$$\begin{aligned} [\lambda] &= \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ -\mu \end{bmatrix} \\ [M] &= \begin{bmatrix} C_{01} \\ C_{02} \\ \vdots \\ C_{0n} \\ 1 \end{bmatrix} \\ [K] &= \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} & 1 \\ C_{21} & C_{22} & \cdots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}. \end{aligned} \quad (15)$$

After solving the weight coefficients λ of the above equations, you can use (8) to calculate and calculate the

predicted values with the valuation points. Since the Kriging interpolation algorithm has a minimum estimation error based on known samples and is considered according to the distribution of the attribute values of the target region, the data of more known sample points within the target region can be used for estimation, and the estimated values are closer to the true value.

We apply the Kriging interpolation algorithm to infer more radio environment information data according to the sample collected information. We use the sample data to estimate the value of sensing area. Our Kriging weights are derived through minimizing the estimator error variance; that is,

$$\min_{\lambda_i \in R} \text{var} (\bar{z}_u^* - \bar{z}_u) \quad (16)$$

under the unbiasedness constraint, given by the following.

$$E [\bar{z}_u^* - \bar{z}_u] = 0 \quad (17)$$

The mathematical expectation of the sample radio environment information is zero. Assuming the intrinsic stationarity and utilizing Lagrange multiplier optimization algorithm to minimize the estimator error variance (16) under the unbiasedness constraint (17), the Kriging weights λ_i in (8) can be calculated as

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ L \end{pmatrix} = \begin{pmatrix} \gamma_{1,1} & \cdots & \gamma_{1,k} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{k,1} & \cdots & \gamma_{k,k} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_{1,u} \\ \vdots \\ \gamma_{k,u} \\ 1 \end{pmatrix} \quad (18)$$

where $\gamma_{i,j}$ is the radio environment information variogram value between the i th and j th neighbor data points, $\gamma_{i,u}$ is the radio environment variogram value between the i th neighbor data point and the interpolation point.

5. Simulation Evaluation

In this section, we show the prototype REM system based on MCS and demonstrate the simulation results of the inference of the missing radio environment information data according to the sample data.

5.1. Implementation of the Prototype System. As is shown in Figure 4, the radio environment information is collected and displayed in the web portal of the platform. As we can see several Wi-Fi properties can be seen on the banner. The properties collected by participants are as follows: SSID, BSSID, frequency, and the Wi-Fi signal strength level. The information belongs to the Wi-Fi signal sources sensed by the nearby participants. The density of the red nodes represents the density of the Wi-Fi signal sources. As we can see in Figure 8 the density of the Wi-Fi signal source is not uniform. This result is as expected.

TABLE 2: Experimental equipment parameter configuration.

AP	MAC	channel	Transmission power (dBm)
AP1	48:7A:DA:B9:21:A0	11	20dBm
AP2	48:7A:DA:B9:12:60	1	20dBm
AP3	50:DA:00:9D:B3:E0	6	20dBm

5.2. Interpolation Performance Evaluation. In this section, experimental settings are described in detail. We will introduce the experiment environment first, and the baseline methods used in the experiments are presented. The experimental data is also introduced, and experimental settings and evaluation metrics are also proposed to evaluate the performance of our method.

5.2.1. Experimental Settings. The experimental area is 150m×150m as the sensing area to collect the radio environment information and infer the missing radio environment information data. In order to simulate the wireless network environment of the target area, three wireless network access points (APs) and radio network controllers (ACs) are deployed to form a WLAN in the target area to cover the target sensing area. The AP and AC are H3C WA4320-ACN and H3C WX3010E. In our WLAN network, the frequency band of the electromagnetic wave transmitted by the AP is 2.4 GHz band. In order to better cover the target area and reduce mutual interference between APs, the three APs use the 1, 6, and 11 channels of the 2.4G band respectively, and the transmission power of AP electromagnetic waves is 20 dBm. All parameters are listed in Table 2.

In order to simulate a small number of users collecting radio environment information in the target area, we mesh the target area. We divide the target area into subareas, each of which has a size of $5m \times 5m$, and then the participants collect the WLAN signal in different subareas using mobile devices to collect Received Signal Strength Indication (RSSI), as shown in Figure 7. The device used by the participants to collect the RSSI value is a Lenovo smart phone (Lenovo A3910e70), and the Wi-Fi analyzer is used to obtain the received signal strength value of the wireless network in the subarea. In order to prove that using the Kriging spatial interpolation algorithm to infer the wireless network environment data in our REM platform has higher accuracy, control experiments are set up. In the control experiments, the missing wireless network environment in the target area is restored by Nearest Neighbor (NN) and Inverse Distance Weighting (IDW) according to the sample data of the wireless network environment data collected by the participants.

5.2.2. Baseline Methods. To verify the high accuracy of the Kriging spatial interpolation algorithm, we used the nearest neighbor (NN) interpolation algorithm, inverse distance weighted (IDW) interpolation algorithm, and Kriging interpolation algorithm to predict the restoration goal under the same data volume of sample data.

(a) NN [21, 22]: Nearest neighbor interpolation is a simple method of multivariate interpolation in one or more

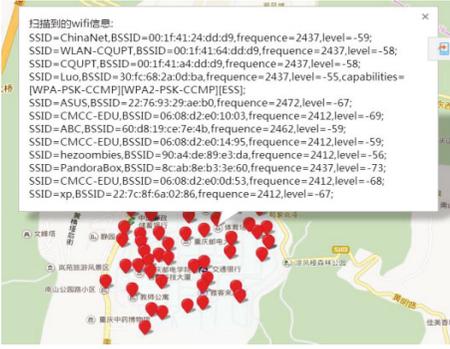


FIGURE 4: The web portal of REM based on MCS.

dimensions. For a given set of points in space, a Voronoi diagram is a decomposition of space into cells, one for each given point, so that anywhere in space, the closest given point is inside the cell. This is equivalent to nearest neighbor interpolation, by assigning the function value at the given point to all the points inside the cell.

(b) IDW [23]: Inverse distance weighting (IDW) is a type of deterministic method for multivariate interpolation with a known scattered set of points. The assigned values to unknown points are calculated with a weighted average of the values available at the known points.

5.2.3. Experimental Data. In order to obtain the discrete data points calculated from the experimental variogram based on the WLAN sample data collected in the target area, a theoretical variogram model and a power function model were selected according to the distribution of the discrete data points. To verify the effect of different numbers of sample data on inference of the radio environment data of the entire target area, the target area is divided into 315 subareas, and a scale factor of the number of sample data and total data is set. During the experiment, the WLAN electromagnetic wave signal propagates in the free space and causes attenuation. Therefore, the smaller the distance from the electromagnetic wave transmission position, the greater the RSSI value that can be obtained. We choose $\rho = 0.1$ and $\rho = 0.3$ to indicate the percentage (10% and 30%) of subareas being covered by the participants of the target area, respectively; by drawing the target area WLAN signal RSSI heat map, we can compare the accuracy of our algorithm under different coverage situations. According to the sample data collected from the sample, the heat map of the RSSI is plotted using the Kriging spatial interpolation algorithm.

Figure 5 presents the empirical variogram and fitting result of a beacon in radio environment information. We input the sample information data to Matlab. The fitted curve demonstrates the spatial correlation model of data and is used to estimate the information data at a sensing location. As shown, the value of empirical variogram, which is the scatter plot in Figure 7, increases with h . It infers that there is an obvious trend (general spatial variation of the mean value) of RSS distribution in the area. Compared with more widely used fitting functions, e.g., the spherical and exponential

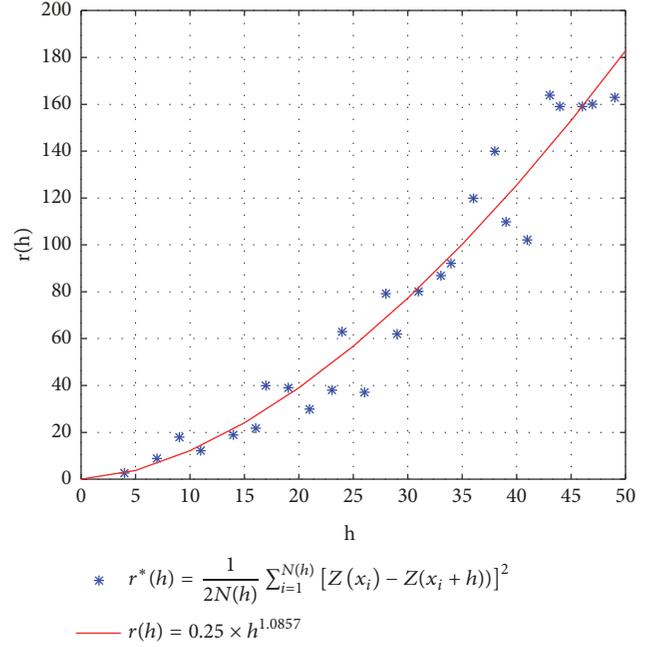


FIGURE 5: Variation function discrete data points and fitting curves.

function, it is suggested that a power model is selected; that is,

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \alpha h^\beta, & h \geq 0 \end{cases} \quad (19)$$

where α and β are the fitting parameters with a strict constraint that $0 < \beta < 2$.

As indicated in the figure, the power function is well fitted, so we choose the power function as the variogram. It can be obtained by fitting with Matlab fitting function, as is shown in Figure 5. After the Matlab fitting function, we can get the variogram being $r(h) = 0.25 \times h^{1.68}$. Then, we can calculate the variogram between all known points. The value of sample RSSI points is related to the distance.

5.2.4. Evaluation Metrics. First, the signal propagation is simulated over the interest area. Meanwhile, the interpolation error of the Kriging interpolation algorithm is compared with nearest neighbor (NN) interpolation algorithm and inverse distance weighted (IDW) interpolation algorithm.

5.3. Prediction Performance Analysis. In this section, the performances of proposed method are evaluated. First, we show the results of user latent interest distribution. Then, the impact of interest number and the proportion of training set on link prediction can be verified.

5.3.1. Signal Propagation. When $\rho = 0.1$, 10% of all subareas of the target area are covered by the participants to collect WLAN RSSI values, and then we use all the WLAN data restored by the Kriging spatial interpolation algorithm according to 10% of the sample data of all data in the target

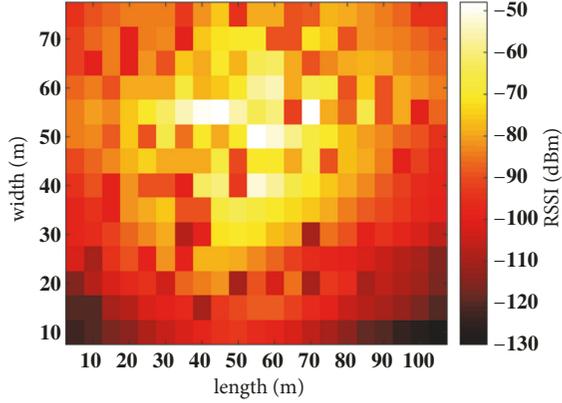


FIGURE 6: When $\rho=0.1$, using the Kriging interpolation algorithm restores the target area WLAN signal RSSI heat map.

area, as shown in Figure 6. The lighter color of the heat map indicates that the signal power of the WLAN received by smart phones at the location is greater, whereas the darker the heat map color, the smaller the WLAN power received by phones at the location. In Figure 6, it can be seen that the color in some areas is not smooth enough, and the RSSI data value fluctuates greatly and its continuity is poor. Figure 7 shows that 30% of all subareas of the target area are covered by the perceived user and collect valid sensory data. Then 30% of all the data in the target area are restored using the Kriging interpolation algorithm to restore the data of all WLANs in the target area. The color in the heat map is excessive. Comparing the two figures, we found that the WLAN RSSI value of the target area restored by spatial interpolation using 30% of all data using the target area is more accurate than the WLAN received power of the restored 10% of the target area using all the data.

5.3.2. Interpolation Algorithm Error Comparison. To verify the accuracy of the Kriging spatial interpolation algorithm, we used the nearest neighbor interpolation (NN) algorithm, inverse distance weighted (IDW) interpolation algorithm, and Kriging interpolation algorithm to predict the restoration goal under the same data volume of sample data.

Figure 8 shows the error comparison of the NN interpolation algorithm, IDW interpolation algorithm, and Kriging interpolation algorithm when the sample data occupies all the data in the target area. To verify the accuracy of different interpolation algorithms, we selected five subareas in the target area as comparison regions interpolated using different interpolation algorithms, and the positions of the five subareas are shown as the positions of the five blue stars in the target area, shown in Figure 8. Interpolation error calculation process is to choose to select each of the five subareas to use three different interpolation algorithms to calculate the estimated value of the point based on the nearby sample data, and then take the absolute value of the estimated value and the measured value of the point. Then the experiment is repeated several times to get the average value and the

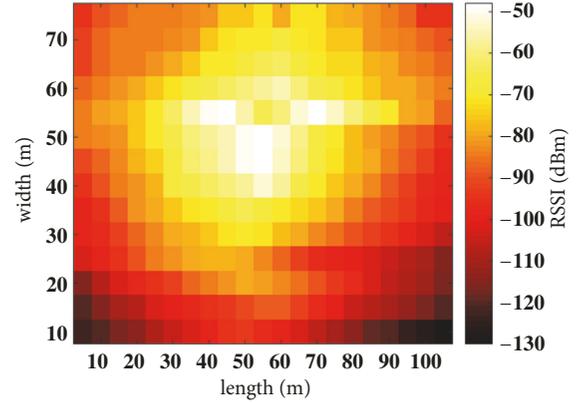


FIGURE 7: When $\rho=0.3$, using the Kriging interpolation algorithm restores the target area WLAN signal RSSI heat map.

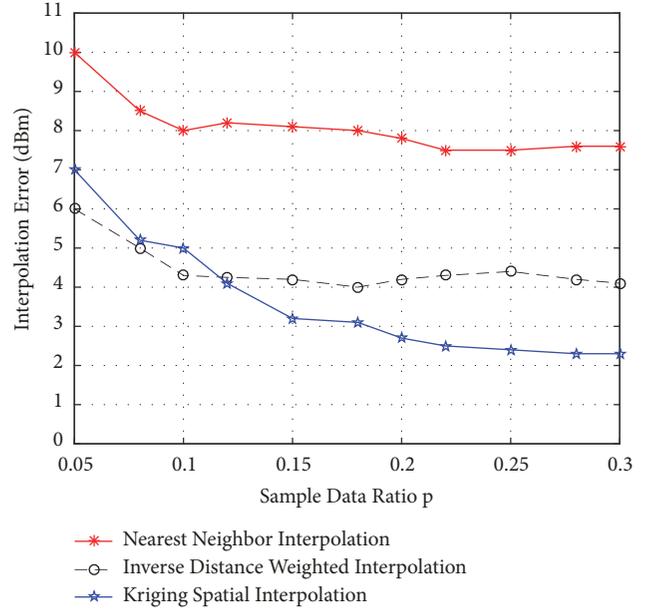


FIGURE 8: The interpolation error comparison of NN interpolation algorithm, IDW interpolation algorithm, and Kriging interpolation algorithm when the sample data proportion is different.

error using different interpolation algorithms, which can be expressed as follows:

$$e = \frac{1}{5n} \sum_{i=1}^n |z_i - z^*|, \quad i = 1, 2, \dots, n \quad (20)$$

where e denotes the error value after one of the interpolation algorithms uses multiple interpolations. n is the number of interpolation experiments using this interpolation algorithm. z_i indicates that this point uses some interpolation algorithm to obtain the estimated value based on the sample point data near the location. z^* is the measured data in the target area.

Figure 8 shows the error of different interpolation algorithms when the sample data occupies different proportions of the overall data. When the sample data occupies 0.05 of the total data, the interpolation error of the three difference

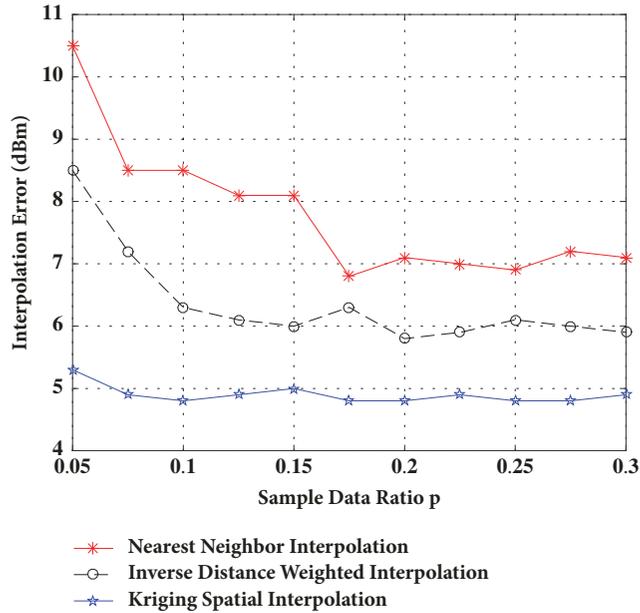


FIGURE 9: The interpolation error comparison of NN interpolation algorithm, IDW interpolation algorithm, and Kriging interpolation algorithm when the sample data proportion is different.

algorithms is relatively large. The interpolation errors of the nearest neighbor interpolation, Kriging spatial interpolation and inverse distance weighting are 10dBm, 7dBm, and 6dBm, respectively. With the increase of the proportion of sample data, the errors of the three interpolation algorithms are reduced. Among them, the error of Kriging spatial interpolation algorithm decreases sharply. Then the error of three interpolation algorithms tends to be stable, and it can be clearly seen that when the nearest neighbor interpolation algorithm is used, interpolation has the greatest error. Moreover, the IDW interpolation algorithm and Kriging spatial interpolation algorithm have smaller error when restoring the WLAN electromagnetic wave environment of the target area according to the sample data.

Figure 9 shows the values of the radio environment data obtained by interpolating sample data from the nearest five regions in a subregion of the target region using the nearest neighbor interpolation algorithm, inverse distance weighted interpolation algorithm, and Kriging interpolation algorithm. It can be seen from the figure that as the proportion of sample data increases; that is, the number of data samples increases; the errors obtained by using the three interpolation algorithms and the real data first decrease. Moreover, the NN interpolation algorithm has the largest error variation when the sample data is less than 0.2, the interpolation error of the IDW interpolation algorithm decreases as the sample data increases.

6. Conclusion

In this paper, we first introduced MCS to collect data for REM construction and proposed a system architecture to collect the radio environment information. Limited by the

budget, only some of the participants can join the sensing task to collect the radio environment information, which leads to incomplete radio environment information. To solve the problem, the Kriging algorithm is proposed to infer the missing radio environment information data with collected sample radio environment information data. The performance is compared to the NN and IDW algorithms over different levels of sparsity. The simulation result shows that Kriging interpolation algorithm can infer the missing radio environment information data and generates more accurate radio environment information data than the NN and IDW algorithm. In the future, some data estimating and processing methods [24–26] can be used to add the sensing data for constructing the radio environment map.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was jointly supported by (1) National Natural Science Foundation of China (No. 61771068, 61671079) and (2) Beijing Municipal Natural Science Foundation (No. 4182041).

References

- [1] P. Singh, M. Kumar, and A. Das, “Effective frequency planning to achieve improved KPIs, TCH and SDCCH drops for a real GSM cellular network,” in *Proceedings of the 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT)*, pp. 673–679, India, July 2014.
- [2] Z. Hou, Y. Zhou, L. Tian, J. Shi, Y. Li, and B. Vucetic, “Radio environment map-aided doppler shift estimation in LTE railway,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4462–4467, 2017.
- [3] K. Ichikawa and T. Fujii, “Radio environment map construction using Hidden Markov Model in multiple primary user environment,” in *Proceedings of the 2017 International Conference on Computing, Networking and Communications, ICNC 2017*, pp. 272–276, January 2017.
- [4] K. Katagiri, K. Sato, and T. Fujii, “Crowdsourcing-Assisted Radio Environment Maps for V2V Communication Systems,” in *Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pp. 1–5, September 2017.
- [5] R. K. Ganti, F. Ye, and H. Lei, “Mobile crowdsensing: current state and future challenges,” *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.
- [6] B. Guo, Z. Wang, Z. Yu et al., “Mobile crowd sensing and computing: the review of an emerging human-powered sensing paradigm,” *ACM Computing Surveys*, vol. 48, no. 1, article 7, 2015.

- [7] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 29–35, 2014.
- [8] J. Riihijarvi, J. Nasreddine, and P. Mahonen, "Demonstrating radio environment map construction from massive data sets," in *Proceedings of the 2012 IEEE International Symposium on Dynamic Spectrum Access Networks, DYSPAN 2012*, pp. 266–267, October 2012.
- [9] S. Ulaganathan, D. Deschrijver, M. Pakparvar et al., "Building accurate radio environment maps from multi-fidelity spectrum sensing data," *Wireless Networks*, vol. 22, no. 8, pp. 2551–2562, 2016.
- [10] S. Ureten, A. Yongacoglu, and E. Petriu, "A comparison of interference cartography generation techniques in cognitive radio networks," in *Proceedings of the ICC 2012 - 2012 IEEE International Conference on Communications*, pp. 1879–1883, June 2012.
- [11] J. Liang, M. Liu, and X. Kui, "A survey of coverage problems in wireless sensor networks," *Sensors & Transducers*, vol. 163, no. 1, pp. 240–246, 2014.
- [12] Z. Wei, Q. Zhang, Z. Feng, W. Li, and T. A. Gulliver, "On the construction of Radio Environment Maps for Cognitive Radio Networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference, WCNC 2013*, vol. 1, pp. 4504–4509, April 2013.
- [13] V. Atanasovski, J. Van De Beek, A. Dejonghe et al., "Constructing radio environment maps with heterogeneous spectrum sensors," in *Proceedings of the IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN 2011*, vol. 3, pp. 660–661, May 2011.
- [14] M. Srivastava, T. Abdelzaher, and B. Szymanski, "Human-centric sensing," *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 370, no. 1958, pp. 176–197, 2012.
- [15] A. A. Khan, M. H. Rehmani, and A. Rachedi, "Cognitive-Radio-Based Internet of Things: Applications, Architectures, Spectrum Related Functionalities, and Future Research Directions," *IEEE Wireless Communications Magazine*, vol. 24, no. 3, pp. 17–25, 2017.
- [16] A. Yaqot and P. A. Hoeher, "Efficient Resource Allocation in Cognitive Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6349–6361, 2017.
- [17] J. Talvitie, M. Renfors, and E. S. Lohan, "Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1340–1353, 2015.
- [18] S. Grimoud, B. Sayrac, S. Ben Jemaa, and E. Moulines, "An algorithm for fast REM construction," in *Proceedings of the 2011 6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications, CROWNCOM 2011*, pp. 251–255, June 2011.
- [19] A. Umbert, F. Casadevall, and E. G. Rodriguez, "An outdoor TV band Radio Environment Map for a Manhattan like layout," in *Proceedings of the 13th International Symposium on Wireless Communication Systems, ISWCS 2016*, pp. 399–403, September 2016.
- [20] D. Zhang, L. Wang, H. Xiong, and B. Guo, "4W1H in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 42–48, 2014.
- [21] R. Olivier and C. Hanqiang, "Nearest Neighbor Value Interpolation," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 4, 2012.
- [22] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.
- [23] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the ACM 23rd National Conference (ACM '68)*, pp. 517–524, 1968.
- [24] Z. Ma, J. Xie, H. Li et al., "The role of data analysis in the development of intelligent energy networks," *IEEE Network*, vol. 31, no. 5, pp. 88–95, 2017.
- [25] S. P. Bharati, F. Cen, A. Sharda, and G. Wang, "RES-Q: Robust Outlier Detection Algorithm for Fundamental Matrix Estimation," *IEEE Access*, vol. 6, pp. 48664–48674, 2018.
- [26] Z. Ma, J.-H. Xue, A. Leijon, Z.-H. Tan, Z. Yang, and J. Guo, "Decorrelation of neutral vector variables: theory and applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 129–143, 2018.

