Reconstruct Radio Map with Automatically Constructed Gaussian Process for Localization

***Abstract*—Over the past decade, the increasing use of wireless networking technology has been implemented in real-world location-based application, especially in localizing clients holding mobile devices in indoor and outdoor environments. According to recent published literature, fingerprint methods outperform other methods and techniques (time-based, angle based, and power-based techniques) in indoor positioning scenarios. Also, many algorithms have been proposed to model radio maps during offline phase and utilize these maps to predict a mobile client's location from received signal strength indicators values obtained from multiple access points. These technologies can be applied to outdoor environment as well. We cooperate with HUAWEI and collect the RSRP(Reference Signal Receiving Power) on main roads and utilize Gaussian Process Regression to predict the values on small roads to construct the radio map in the city for localization. The main contribution of this paper are as follows. First, we identify the different causes of variation in wireless channel quality. Second, we introduce a Bayesian regression model - Gaussian Process Regression model to profile the RSRP pattern of the base stations. We compose kernels automatically and construct the model. Third, we select different models to do the comparison and model ensembling. Later, we present our evaluation in Shanghai JiaoTong University and Yindu Road for 3G and 4G radio map recovery. We collect one million sample points from twenty square kilometers area and get about 5% and 7% error rate for 3G and 4G, respectively.**

I. Introduction

Over the past decade, the increasing use of wireless networking technology has been implemented in real-world location-based applications, especially in localizing clients holding mobile devices in indoor environments. Though GPS is widely used in outdoors, it also has some disadvantages, such as power consuming and time consuming. The most significant disadvantage is that the GPS module must be turned on by the user, or there is no way to locate a device. The situation that we need to locate a device without the user's help exists when there is an emergency such as '911' and kidnapping. So, an accurate alternative wireless positioning for outdoors is needed. Fortunately, there have been many solutions proposed to locate indoor mobile clients due to widespread deployment of WiFi and these technologies can be applied to outdoors as well.

Most adopted solutions for indoor localization use *fingerprinting* of ambient environment signatures. At the same time, a more successful technique requires a substantial 'predeployment' effort by way of creating a *radio map*. Over the last decade, many algorithms have been proposed to model WiFi radio maps and use this to predict a mobile client's location from Received Signal Strength Indicators (RSSI) values obtained from multiple access points (APs).

Fingerprint wireless localization techniques is performed on two phases; offline phase and online phase. During the offline phase, the signal strength received from the access points at selected locations are stored in a database, resulting in a radio map. During the online phase, the system use the signal strength samples of mobile clients from the access points to search radio map to estimate the client's location. In this paper, we develops a new probabilistic techniques based localization system that increases the accuracy of positioning.

The biggest challenges with fingerprints techniques is that they are vulnerable to several dynamic factors in candidate locations. The first kind of factors are *inherent* factors, such as different candidate locations and changes of wireless channels over time when the client is standing at a fixed location. Another kind of factors are *external* factors, such as WiFi hardware variance problem: the WiFi device (training device) used to contrust radio maps during the offline phase differ from the WiFi devices (positioning device) used during the online phase, or difference between positioning devices. To handle such factors it is imperative to first understand the impact of these factors. The experiments described later indicate how these factors influence the accuracy of localization. Further, we also consider these factors into our localization model.

Second, for a large outdoor environment, it is too difficult to sample thousands of survey sites to construct a fine grain radio map. For example, a university usually needs a hundred thousand training data during offline phase. Hence, another objective we investigate in this paper is: *How to deal with large problems when the construction process highly scales and it is possible to reduce the data matrix rank to approximate the radio map, while keeping localization accuracy within a certain limit?* A possible solution is that we can easily collect the data from the main road by driving a car with a mobile device along these roads. The small roads are hard to access and time-consuming to go through. Alone these lines, we introduce machine learning to solve this problem. We use main road data to recover the data on small road and then construct the whole radio map.

The main contributions of this paper are as follows.

- We identify the different causes of variation in the indoor wireless channel quality. The main targets we measure are received signal strength and throughput. The measurement results confirm that the distribution of received signal strength around an AP is Gaussian given the target device is stationary at a location.
- We introduce a Bayesian regression model - Gaussian Process Regression model to profile the RSRP pattern of

the base stations, constructing radio maps. GPR model has been recently used to solve complex machine learning problems.The GPR is specified by its mean function and covariance function. We use composing kernels to do the automatic model construction to get better accuracy.
- We select different models such Linear Regression, Support Vector Regression, Gradient Boost Tree, Xgboost and so on to do the model comparison and model ensembling and evaluate their performance.

The remainder of the paper is organized as follows. We introduce the math background of Gaussian Process in Section III.We then introduce wireless channel characteristic measurement experiments in Section IV. We describe our Gaussian Process model in Section V and give automatically kernel composing. In section VII, we present our evaluation. Related work and conclusions are in Section II and Section VIII, respectively.

II. Related Work

Localization has been an active area of research for the past two decades. In the following, we highlight some of the work on localization in wireless systems and the scheme on radio map reconstruction, and point out the difference in our work.

There are two popular techniques for indoor localization based on wireless network, triangulation-based method and RF-based method. In triangulation-based method, it uses the geometric properties of triangles to estimate the location of the target. Triangulation-based technology can be mainly categorize into two group: lateration and angulation. In lateration, it estimates the location by measuring its distances between target and the multiple reference points. Such technology include TOA (time of arrival)–using different signaling techniques such as direct sequence spread-spectrum (DSSS) [2], [3] or ultrawide band (UWB) measurements [1], [4] – and TDOA (time difference of arrival) [5]. In angulation, it estimate the location by computing angles retative to multiple reference points.One of the technology used angulation is AOA Estimation [8].

Comparing with the triangulation-based technology, the RF-based technology does not need additional hardware but a WiFi-integrated mobile device.RF-based scene analysis refers to the type of algorithms that pre-collect features (fingerprints) of a scene and then estimate the location of an object by matching online measurements with the closest a priori location fingerprints. Two main category of fingerprint-based technology is kNN-based method and probabilistic-based method.kNN (k nearest neighbor algorithm) estimates the target's location by computing the centroid of the k closest neighbors that have the smallest euclidean distance to the online RSS reading data, such work including RADAR by Microsoft. [9], [10]. Based on the survey of Liu [6],the probabilistic method (such that Horus [7]) have better performance than the kNN method. And our work develop a novel probabilistic techniques in order to increase the accuracy.

One problem for fingerprint-based method is that it need significant labour cost during the offline parse to construct the radio map. To solve this problem we use a small number of fingerprint to reconstruct the whole radio map. Feng *et.al.* [11]develop a compressive sensing scheme to reconstruct the radio map based on RSS measurements at only a subset of fingerprints. However, compressive sensing scheme base on the assumption that the radio map have a sparse nature.And this work does not care about the temporal characteristic of RSS.

The other external factor that will influence the accuracy of the localization the device variance problem. Tsui *et.al* [12]'s work focus on the hardware variance problem of RSS-based localization. Their work base on the assumption that there is a linear shift between different devices. In our work we try to use Gaussian Process to find the transformation function, and combine this model to the spatio-temporal model.

III. Background

The core of our method is a Gaussian process regression model that describes the characteristic of radio map. Before discussing the characteristic of radio map and its use for radio map reconstruction, we give a brief overview of Gaussian process and its regression.

Because of the drawbacks of radio propagation model, an alternative method is probability techniques with supervised learning. Supervised learning, that is, given empirical data (the training dataset), we learn the relationship between input and output, then make the prediction based on new observations. In our problem, we take the main road data as the training set and the small road data as the test set. Here, we do not consider a strict function between input and output by violence. On the contrary, we give a prior probability to every possible function, where higher probabilities are give to functions that we consider to be more likely. That is what Gaussian process is going to do. Here, we don't rigidly differentiate *process* and *distribution*, though a probability distribution and a stochastic process respectively describes random variables and properties of functions.

A. Gaussian Process

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.In this section we give a brief review of GPR for implementation purpose; further details can be found in [**?**]. GPML website: http://gaussianprocess.org/gpml/

A Gaussian process $f(\mathbf{x})$ is completely determined by a mean function $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$,such that, where $\mathbf{x}$ is the input vector

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{1}$$

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{2}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \tag{3}$$

Concerning with Gaussian noise, a Gaussian process can be expressed as

$$y = f(\mathbf{x}) + \varepsilon \tag{4}$$

where $X = \{\mathbf{x}_i | i = 1, ..., n\}$ is input dataset, $\mathbf{y} = \{y_i | i = 1, ..., n\}$ is output and $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$.

B. Gaussian Process Regression

The joint distribution of the training outputs $\mathbf{y}$ and the test outputs $\mathbf{f}_*$with a zero mean function is

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}(0, \begin{bmatrix} k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I & k(\mathbf{X}, \mathbf{X}_*) \\ k(\mathbf{X}_*, \mathbf{X}) & k(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}) \tag{5}$$

where $\mathbf{X}$ and $\mathbf{X}_*$ are design matrices for training data and test data respectively.Conditioning $\mathbf{f}_*$ on observation $\mathbf{y}$, the predictive distribution can be derived:

$$p(f_*|\mathbf{x}_*, X, \mathbf{y}) \sim \mathcal{N}(\overline{\mathbf{f}}_*, V(\mathbf{f}_*)) \tag{6}$$

where

$$\overline{\mathbf{f}}_* = k(\mathbf{X}_*, \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1} y, \tag{7}$$

$$V(\mathbf{f}_*) = k(\mathbf{X}_*, \mathbf{X}_*) - k(\mathbf{X}_*, \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1} k(\mathbf{X}, \mathbf{X}_*). \tag{8}$$

$$\begin{aligned} p(f_*|\mathbf{x}_*, X, \mathbf{y}) &= \mathcal{N}(\phi_*^\top \textstyle\sum_p \Phi(K + \sigma_n^2 I)^{-1} \mathbf{y}, \\ &\phi_*^\top \textstyle\sum_p \phi_* - \phi_*^\top \textstyle\sum_p \Phi(K + \sigma_n^2 I)^{-1} \Phi^\top \textstyle\sum_p \phi_*) \end{aligned} \tag{9}$$

Further, we can define $k(\mathbf{X}, \mathbf{X}')$ is a *covariance function or kernel*.

Under normal circumstances, we assumes the Mean function of Gaussian process as zero, because it turns out that the impact is adding an item related to noise variance at the end of the covariance function (kernel) and, so it is equivalent to a new kernel, which will not affect the final result.

IV. Observation

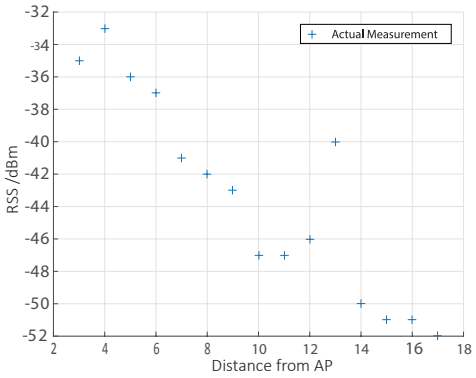
Due to the very dynamic nature of the wireless signal, the assumption that the wireless characteristic during the offline phase keep consistent with the condition during online phase is impractical. Thus, it seems more necessary to monitor RSSI characteristic and variations in both spatial and temporal domains.

A. Spatial Characteristic

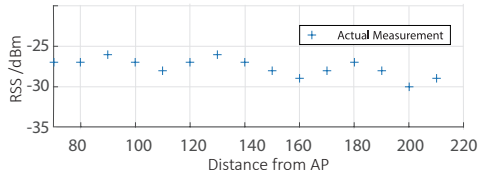
This section describes how the RSS value receiving from one AP changes when the location of detector varies. The variation is divided into two characters, one meaning distance between the detector and the AP and the other meaning detector's orientation towards the AP.

1) Distance variation: We conduct our measurement in the test bed, and measure RSS values from one AP as distance from it increases. The selected test locations align as a line as possible. Because the wifi networks work at the 2.4GHz range, the wavelength is 12.5cm, we measure the RSS values under *two situations*: 1) the distance interval is larger than 12.5 centimeters, specifically, we set interval about 0.5 meters. 2) the distance interval is less than 12.5 centimeters.

Considering the first situation, Fig.1(a)shows the RSS value receive from AP will decrease with the distance increasing. However the relationship between them is *not a strict linear function*. Under the second situation, the RSS value varies in less than 1dBm. Remember our goal is construct the radio map more efficiently and reduce unnecessary squander of labor during offline phase. So, we don't need to capture the the RSS variance in the wavelength range.



(a) Under the short distance situation



(b) Under the long distance situation

Fig. 1. Fluctuation of RSS

2) Orientation variation: These variation happen when the angle between the detector and the AP changes. Theoretically, the attenuation of wireless signal strength is identical over all orientation. However, it is impossible due to specific indoor architectural structure. The position of windows, doors, and corners could all influence the distribution of RSS values [**?**,]. It should be noted that some work have studied that people standing at different angles to an AP (facing north or south, but staying at the same site) may cause different attenuation to the radio wave power [**?**,]. In our opinion, the influence coming from the variation of human body is less significant than the influence by specific spatial structure. Obviously, the latter plays a pivotal role in practical situation.

B. Temporal Characteristic

In this section, We mainly focus on the temporal changing property of wireless signal strength. In order to observe this property, we settle the mobive devices in a fixed place and let them collect the wireless signal strength from a given access point. An Android application we programmed is used to collect the signal strength data for a long period of time. The smart phone used in this experiment is the Note-1s from Xiaomi company, and the access point is provided by Foxcoon company. This AP works on dual band, which 2.4GHz and 5 GHz. In our experiment,the frequency band we used is 2.4GHz. We collect data for two hours in three period of time: 10:00 12:00, 16:00 18:00 and 20:00 24:00. The result of our experiment is shown as following figures.

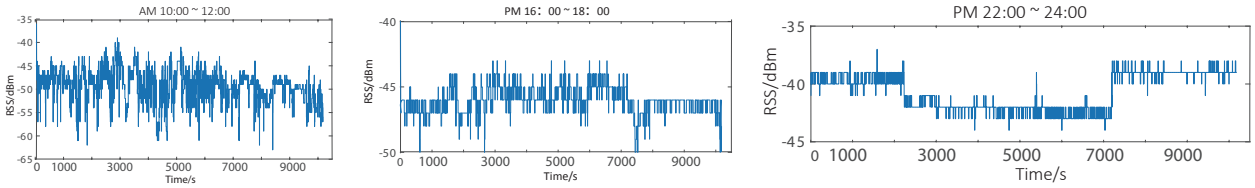


Fig. 2. Fluctuation of RSS

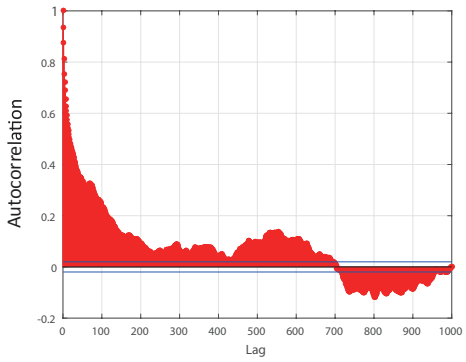


Fig. 3. Autocorrelation of temporal sequence of RSS

From the Fig.2, we can exploit that the fluctuation of the RSS is varied in different time period. The reason of the variety is partily because the activity of the client conecting the access point can influence the RSS value smart phone collected.

In order to mine the information behind the data, we analyse the autocorrelation of the temporal RSS sequence. The result is showen as following figures.

From Fig.3, we can get that the autocorrelation is big when it has a short delay, which indicates that there is a strong correlation between the sequence in a short period of time. In conclusion, the prediction of temporal RSS is practicable, since if there is a correlation in a short period of time, the RSS in this period of time can be used in the prediction of the following time period. We can also find out that if the delay time is long, the autocorrelation will rebound, however, the autocorrelation is still a small value. This observation shows that the appearance of sequential wave crest does not indicate the periodicity of temporal RSS sequence.

V. Kernel Composing

In the traditional offline phase, a site survey performed in the selected area is conducted by an operator who collects RSRP values at target locations to construct the database as a function of the client's physical coordinates. The drawback of this method is time-consuming, especially impractical under large outdoor environments. In section IV-A and section IV-B, we have manifested that existing RSRP patterns for each base station. Therefore, an alternative approach is establishing a power profile for each base station, utilizing small number of RSRP observations to estimate or predict RSRP values at unknown locations.

We propose the following regression model based on statistical knowledge. By training a small number of data points, we can give a reasonable estimate. The Gaussian process regression model we used has described in Sec. III.

A. kernel methods

Suppose our training dataset is $\mathcal{D} = \{X, \mathbf{y}\}$. Next, we will specify the input and output of Gaussian process used in our approach.

Let us consider a two-dimensional physical space $\mathbb{X}$, with n base stations. Due to most base stations are previously installed into the environment, we just assume the location of base stations is known. Let us denote the RSRP reported from various base stations by an n dimensional vector $\mathbf{s} \in \mathbb{S}$, where $\mathbb{S}$ is the n-dimensional signal strength space. We select m locations as fundamental dataset. Hence, at each location i, we obtain RSRP values vector $L_i = [rss_1, rss_2, ..., rss_n], i = 1, 2, ..., m$ reported from n base stations. As for each AP installed into $\mathbb{X}$, we can also write the dataset into the form of $M_j = [rss_1, rss_2, ..., rss_m]^\top, j = 1, 2, ..., n$ reported from m locations. Therefore, the fundamental dataset from offline observations (basic radio map) is represented by $\Psi^{m \times n}$ where $\psi_{i,j} = \frac{1}{t}\sum_{\tau=1}^{t} \psi_{i,j}(\tau)$ is the average of RSS readings over time domain from base station j at location i, as $\{\psi_{i,j}(\tau), \tau = 1, ..., t, t > 1\}$ with t being the total number of time samples collected. Because of limited power of one base station, there is might no RSS readings for an base station at some locations. The corresponding RSS values in the Ψ are set to a very small values (-110dBm in our approach) which implies zero power readings. One thing need to note is that in our approach, the dynamical phenomenon in the target space $\mathbb{X}$ is specified by a spatio-temporal Gaussian process, hence t is supposed to be small when we need to observe RSS values distribution over long time $\mathcal{T}$. Here t is just used to eliminate the measurement error, and acuqiring average readings could reduce this error.

In the next subsection, we will specifically discuss how to use Gaussian process to estimate the RSRP values from various base stations at unknown locations.

The kernel function(covariance function) plays the important role in Guassian Process Regression as it corresponds to a kind of model assumption. We note that each kernel function has a series of parameters that can be used to control the specific shape of the covariance function. Here some parameters are called *hyperparameters* because *hyperparameters* do not

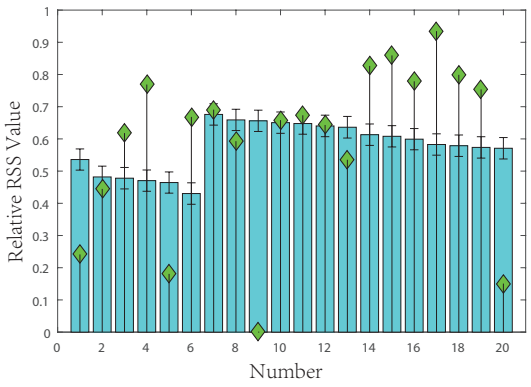


Fig. 4. The result of using SE kernel directly

directly determine the form of the function, but the distribution of other parameters.

According to the test results of Section IV-A, there is not direct periodic characteristic and linear characteristic. So choosing Periodic kernel function or Linear kernel function is unwarranted.So we model the radio map as a Guassian Process and choose square exponential kernel function:

$$k(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp(-\sum_{l\in x,y} \frac{(s_l - s_l')^2}{2\ell^2}) \tag{10}$$

where $(s_x, s_y) = \mathbf{x}$ represents a coordinate of two-dimensional physical space $\mathbb{X}$. ℓ and σ_f^2 is the length-scale and the signal variance respectively. Generally, we call these parameters *hyperparameters* to emphasize that they are parameters of non-parametric model.

Figure 4 shows the predict result of using SE kernel directly. We selected part of the location data as training set, and then predict the wireless signal strength value in the rest of the position, which are called test points. Here we do not consider the influence of time. We list the test points in the order from small to large in x axis, and the wireless signal strength value in the y axis. We draw the true value in the picture. The predicted value is indicated by the mean and variance of the model output. The upper bound of the prediction equals to the mean plus twice the standard deviation, the lower bound of the prediction equals to mean minus twice the standard deviation. As you can see,there is still a large gap between the real and predicted values, so the prediction did not meet our expectation.

Here we need to build more complex Gaussian kernel functions. To construct combined kernel functions, there are two ways, one is Addition, the other is Multiplication.

(1)

$$\mathcal{K}_a + \mathcal{K}_b = \mathcal{K}_a(x,x') + \mathcal{K}_b(x,x') \tag{11}$$

The kernel function constructed by Addition has all the properties of the original basic functions. For example,the addition of periodic and linear kernel functions shows linear trend as a whole, but they are approximately periodic in a small area. Taking advantage of this feature, we can construct a new kernel function in space.

Considering the measurement in the second chapter, we use two SE kernels, which have different parameters in length scale. We construct such a new kernel function because the wireless signal strength in the area close to the Access Point changes more intense than those areas far away from the AP, which fits the test above:more than fourteen meters away, the change of the RSS value is not that obvious.

So we construct a new Gaussian kernel function as:

$$k(\mathbf{x},\mathbf{x}') = \sigma_{f1}^2 \exp(\frac{(x_l - x_l')^2}{2\ell_1^2}) + \sigma_{f2}^2 \exp(\frac{(x_l - x_l')^2}{2\ell_2^2}) \tag{12}$$

Note that in the formula, two parameters σ^2 is not the same.

So far, this kernel function based on space has a total of four *hyperparameters*. As to how to estimate the *hyperparameters*, there are many methods,commonly used maximizing the probability density methods will be discussed later in detail.

(2)

$$\mathcal{K}_a * \mathcal{K}_b = \mathcal{K}_a(x,x') * \mathcal{K}_b(x,x') \tag{13}$$

Apart from using the Addition method to construct the kernel function, we can also use the Multiplicity method, whose purpose, typically,is to construct a multi-input Gaussian kernel. In particular, the multiplicity of several square exponential kernel function has a special name "Automatic Relevance Determination"(ARD), where the length-scale parameter ℓ_1,ℓ_2 ... determines the correlation between the different dimensions. Those dimensions with larger length-scale means a relatively small change in that dimension. Therefore, we continue rewriting to the Gaussian kernel above as:

$$k(\mathbf{x},\mathbf{x}') = \sigma_{f1}^2 \exp(-\sum_{l\in x,y} \frac{(x_l - x_l')^2}{2\ell_1^2}) + \sigma_{f2}^2 \exp(-\sum_{l\in x,y} \frac{(x_l - x_l')^2}{2\ell_2^2}) \tag{14}$$

So far, this kernel function based on space has a total of eight *hyperparameters*. So far,we did not take time domain into consideration. A simple idea is to multiply the kernel of time dimension to the kernel of space, which introduces too many parameters and costs much time in the parameter estimation. Our model proposes a more reasonable solution and to reduce the computational complexity. Ultimately we can predict the signal strength value in a new time at a new location, so that we can provide a robust signal strength map, reduce the time required and the number of samples.

B. Automatic Construction through Kernel Composing

To automatically construct Gaussian process models, we search over sums and products of kernels, maximizing the approximate marginal likelihood. We show how any model in this class can be automatically decomposed into different parts, and illustrate the structure discovered in the data.

We have used the measured data in the Section **??** to construct the kernel function of the kernel function based on

space. Gaussian process models use a kernel to define the covariance between any two function values: $Cov(y, y') = k(x, x')$. Commonly used kernels families include the squared exponential (SE), periodic (Per), linear (Lin), and rational quadratic (RQ). Positive semi-definite kernels (i.e. those which define valid covariance functions) are closed under addition and multiplication. This allows one to create richly structured and interpretable kernels from well understood base components.

C. Searching over structures

Despite its importance, choosing the structural form of the kernel in nonparametric regression remains a black art. We define a space of kernel structures which are built compositionally by adding and multiplying a small number of base kernels. We present a method for searching over this space of structures which mirrors the scientific discovery process. The learned structures can often decompose functions into interpretable components and enable long-range extrapolation on time-series datasets. Our structure search method outperforms many widely used kernels and kernel combination methods on a variety of prediction tasks.

As discussed above, we can construct a wide variety of kernel structures compositionally by adding and multiplying a small number of base kernels. In particular, we consider the four base kernel families discussedabove: SE, Per, Lin, and RQ. Any algebraic expression combining these kernels using the operations $+$ and $\times$ defines a kernel family, whose parameters are the concatenation of the parameters for the base kernel families.

To evaluate a kernel family we must integrate over kernel parameters. We approximate this intractable integral with the Bayesian information criterion(Schwarz, 1978) after first optimizing to find the maximum-likelihood kernel parameters.

$$\mathcal{BCI}(M) = -2 \log P(D|M) + |M| \log n \qquad (15)$$

where $|M|$ is the number of kernel parameters, $p(D|M)$ is the marginal likelihood of the data, D, and n is the number of data points. BIC trades off model fit and complexity.

VI. Model Ensembling

The Gaussian Prosess Regression is a robust model with high accuracy to do the machine learning. However, we still struggle to get better accuracy in this task. In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. So it is a very powerful technique to increase accuracy on a variety of ML tasks. In this part we will share the ensembling approaches.

Some common advantages of ensemble methods including:1)They average out biases. If you average a bunch of democratic-leaning polls and a bunch of republican-leaning polls together, you will get on average something that isn't leaning either way.2)They reduce the variance. The aggregate opinion of a bunch of models is less noisy than the single opinion of one of the models. This is also why the models will be better with more data points rather than fewer. 3)They're unlikely to overfit. If you have individual models that didn't overfit, and you're combining the predictions from each model in a simple way (average, weighted average, or logistic regression), then there's no room for overfitting.

A. Selected Models

We choose several widely used machine learning methods to do the ensembling: Linear Regression(LR),Support Vector Regression(SVR),Gradient Tree Boosting(GDBT),Xgboost,GP (SE),GP (RQ),GP (Compose). Among them, LR is a model with high bias but low variance. GDBT and Xgboost are already ensemble models and are widely used in all kinds of data mining competitions and the industry as well. GP(SE) and GP(RQ) are traditional GP models. GP(Compose) is the model constructed by automatically kernel composing and selection. As shown in the results, boost methods and nonparametric GP yields better results than the other vector based statistic learning methods.

B. Ensembling Methods

1) Averaging: Averaging works well for a wide range of problems (both classification and regression) and metrics (AUC, squared error or logarithmic loss).There is not much more to averaging than taking the mean of individual model predictions. Averaging predictions often reduces overfitting. We ideally want a smooth separation between classes, and a single model's predictions can be a little rough around the edges. Remember the goal is not to memorize the training data, but to generalize well to new unseen data.

Rank averaging. When averaging the outputs from multiple different models some problems can pop up. Not all predictors are perfectly calibrated: they may be over- or under confident when predicting a low or high probability. Or the predictions clutter around a certain range. Our solution is to first turn the predictions into ranks, then averaging these ranks. After normalizing the averaged ranks between 0 and 1 you are sure to get an even distribution in your predictions.

2) Stacked and Blending: Stacked generalization was introduced by Wolpert in a 1992 paper, 2 years before the seminal Breiman paper 'Bagging Predictors'. The basic idea behind stacked generalization is to use a pool of base classifiers, then using another classifier to combine their predictions, with the aim of reducing the generalization error. Lets say you want to do 2-fold stacking: Split the train set in 2 parts: train_a and train_b. Fit a first-stage model on train_a and create predictions for train_b. Fit the same model on train_b and create predictions for train_a. Finally fit the model on the entire train set and create predictions for the test set. Now train a second-stage stacker model on the probabilities from the first-stage model(s). A stacker model gets more information on the problem space by using the first-stage predictions as features, than if it was trained in isolation.

Blending. Blending is a word introduced by the Netflix winners. It is very close to stacked generalization, but a bit

simpler and less risk of an information leak. With blending, instead of creating out-of-fold predictions for the train set, we create a small holdout set of say 10% of the train set. Blending has a few benefits: It is simpler than stacking. It wards against an information leak: The generalizers and stackers use different data. We do not need to share a seed for stratified folds with your teammates. Anyone can throw models in the 'blender' and the blender decides if it wants to keep that model or not. The cons are: You use less data overall The final model may overfit to the holdout set. The CV is more solid with stacking (calculated over more folds) than using a single small holdout set. As for performance, both techniques are able to give similar results.

We use both rank averaging and blending to do the model ensemble. In the rank average, we rank the models according to their accuracies on the training data. In the blending, we split the test set and use 10% as the training set. Here is a little difference. We use the data from the test set because the data can be sampled in the training phase as well and the environment factors such as the weather and the traffic that day are important as well. We can take these factors into consideration by doing so. The performance of model ensemble is evaluated below.

VII. Performance Evaluation

A. Experiment

In this section, we measure the RSRP values, operating over 3G and 4G in Shanghai JiaoTong University and Yindu Road. We collect one million sample points in about twenty square kilometers. Here we use the main road data as train set and the small road data as the test set(see Fig.6).

1) Experimental Test Bed: We perform our outdoor RSRP measurement in two testbeds used for experiments. The test environment, henceforth called ROE, is the whole campus of Shanghai Jiaotong University and the Zizhu Science Park. The ROE consists of obstructions in the form of buildings roads and lakes.The environments are equipped by base stations are distributed in environments (see Fig.6). There are more than 100 base stations covering the ROE that is the testbed in Section IV for characteristic analysis of RSRP.

2) Experiment on kernel function in time domain: The Rational Quadratic kernel function take both long and short period of time into consideration. We focus on one selected point in the data, and measure the received signal strength from one base station for one continuous minute and model them using the Rational Quadratic kernel functionas shown in Figure 7.

In the figure,the area after the blue line represents the time after the training data, which is the time in the future. Figure **??** is actually made up of two parts, Figure **??** represents the short time,and Figure **??** represents the long time. As can be seen, in a short time, the variation of the radio signal strength is much more complex, for the curve is sharp. On the other hand,in the long time,the wireless signal strength changes more gently, showing relatively smooth curve.

Fig. 5. Floor plan for the testbeds ROE

Fig. 6. Main road as training set and small road as test set

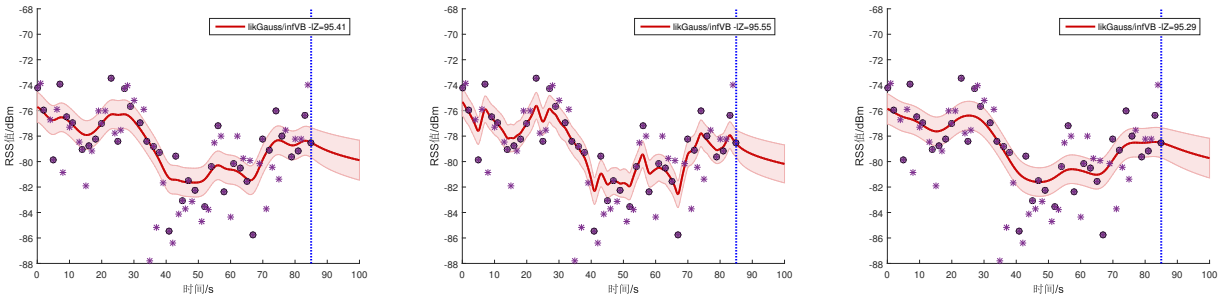
B. Experiment with real data for recovery

1) outdoor environment: For outdoor environment, it is not as complex as the indoor environment, for there is not much influence brought by obstructions such as walls and people.

We get the data provided by HUAWEI company and by ourselves, which are collected by real base stations located in Shanghai. For every base stations, we have from dozens to one thousand measured points and the data covers more than two hundred base stations.

2) recovery process: We use main road data as training data to recover the rest small road data as test set to test whether our Gaussian Process based recovery has a good result.

We take two APs as example: one has 200 measure points and another has 400 points. Using Gaussian Process Regression, we can predict the 20% points after training according to the 80% points. In Figure VII-A2 and Figure VII-A2 we plot

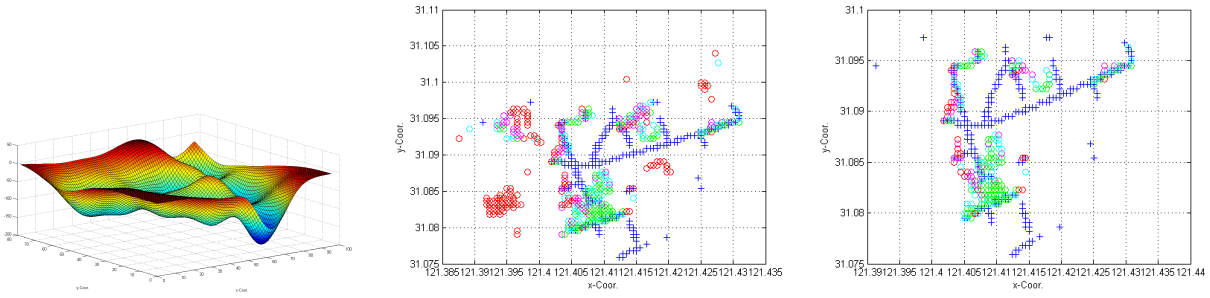


(a) the change of signal strength in a period of time

(b) when l is relatively large

(c) when l is relatively small

Fig. 7. Floor plan for the two testbeds where RSS readings were collected in the marked points. (Black points are building shores.)

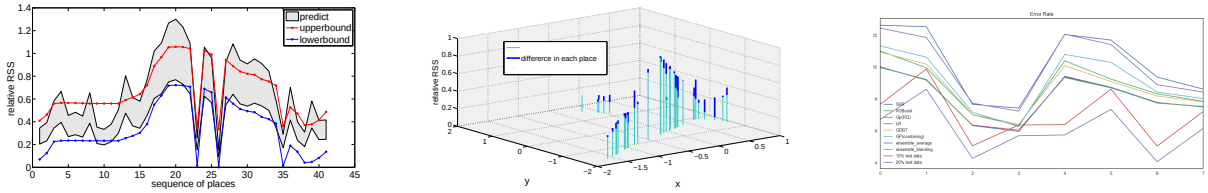


(a) The prediction in whole area.The deeper the color , the larger the RSRP value.

(b) The blue cross as training and others as test.Green: less than 5db, Pink: between 5db and 10 db. Red: more than 10 db.

(c) After deleting points with high variance

Fig. 8. One base station as example



(a) Part of one GP. The gray area indicates the variance

(b) Error in blue and truth is green in the test area, which is small road

(c) Error Rate of different Models

the mean function plus/minus two standard deviations, which corresponds to a 95% confidence interval.

The mean error is less than 5 db, and the standard error is less than 10db. The prediction has good results.

Here gives the meaning of above three figures in Fig.VII-A2. The first one is the recovery result in the whole area. The deeper the color , the larger the RSRP value. The second and the third figure gives the result of this recovery. The blue points are the training data from the main road. The other cross points are the test data in the small roads. The green cross means the error is less than 5db, the pink cross means the error is between 5db and 10 db. The red cross means the error is more than 10 db. By deleting the points with large variance, we can avoid much red points with more than 10db error.

3) Model Comparison: We propose several models above and here we want to give the comparison and evaluate the results(see Fig.VII-B2).

We notice that in single models, boost method such as GDBT and Gaussian Process yields better accuracy than the traditional statistic learning methods. The ensemble method gets better results than all single model, which validates the statement that ensembling can reduce the overfitting problem. With 10% samll road data as training data, we can get about 5% error rate in 4G environment and about 7% error rate in 3G environment.

VIII. Conclusion and Future Work

This paper propose an efficient way to recovery the radio map that provide a foundation for localization in online phase. This method based on signal strength and the theoretical

Unnamed: 0	SVR	XGBoost	Gp(RQ)	LR	GDBT	GP(combining)	nsemble_averag	nsemble_blendir	10% small road	20% small road
SJTU_3G…	12.6	11	10	12.4	10.9	11.3	9.98	10	7.67	6.72
SJTU_3G…	12.5	9.96	9.19	11.9	10.2	10.6	9.23	9.19	9.88	8.62
SJTU_4G…	7.67	7.02	6.37	7.74	7.04	7.18	6.34	6.36	5.07	4.28
SJTU_4G…	7.43	6.34	6.05	7.24	6.42	6.27	5.97	nan	6.37	5.7
YINDU_3…	12.1	10.4	9.38	12.1	10.1	10.8	9.43	9.34	6.4	5.75
YINDU_3…	11.7	9.25	8.72	11.4	9.1	10.3	8.74	8.7	8.62	7.35
YINDU_4…	9.38	8.33	7.75	8.9	8.19	8.44	7.79	7.77	5.05	4.07
YINDU_4…	8.62	7.86	7.53	8.44	7.82	8.03	7.5	7.53	7.22	6.16

(d) Error Rate of different Models(%)

basis is Gaussian process. After taking these two domains into account, we propose a GP automatic kernel construction model, which aims to combine and search for the best kernels and predict the values on the small road so that we can refresh and reconstruct the radio maps. Because of the problem of overfitting, we then propose the model ensembling, that is, by combining different models with averaging or blending, we can obtain a better model. We also carried out extensive experiments to validate our model. Our future work is to reduce the complexity of solving mixed Gaussian process model and taking some domain knowledge into consideration.

REFERENCES

[1] Catarina C. Cruz, Jorge R. Costa and Carlos A. Fernandes," Hybrid UHF/UWB antenna for passive indoor identification and localization systems." *IEEE Transactions on Systems, Antennas and Propagation*,vol. 61,no. 1 ,Jan.2013, pp.1354361.

[2] B. B. Peterson, C. Kmiecik, R. Hartnett, P. M. Thompson, J. Mendoza, and H. Nguyen, Spread spectrum indoor geolocation, *J. Inst. Navigat.*, vol. 45, no. 2, pp. 97102, 1998.

[3] X. Li, K. Pahlavan, M. Latva-aho, and M. Ylianttila, Comparison of indoor geolocation methods in DSSS and OFDM wireless LAN, in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2000, vol. 6, pp. 30153020

[4] N. S. Correal, S. Kyperountas, Q. Shi, and M. Welborn, An ultra wideband relative location system, in *Proc. IEEE Conf. Ultra Wideband Syst. Technol.*, Nov. 2003, pp. 394397

[5] X. Li, K. Pahlavan, M. Latva-aho, and M. Ylianttila, Comparison of indoor geolocation methods in DSSS and OFDM wireless LAN, in Proc. IEEE Veh. Technol. Conf., Sep. 2000, vol. 6, pp. 30153020

[6] H.Liu, H.Darabi, P.Banerjee,J.Liu,"Survey of Wireless Indoor Positioning Techniques and Systems"*IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICSPART C: APPLICATIONS AND REVIEWS*, vol. 37, no. 6, NOVEMBER 2007 1067

[7] M. Youssef and A. Agrawala. "The Horus WLAN Location Determination System." *MobiSys*, 2005.

[8] Niculescu D, Nath B.,"Ad hoc positioning system (APS) using AoA." *Proc. of the IEEE INFOCOM 2003*. Vol.3, 2003. 17341743

[9] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," *Proc. IEEE INFOCOM 2000*, Mar., vol. 2, pp. 775784.

[10] J. Ma, X. Li, X. Tao, and J. Lu, "Cluster Filtered KNN: A WLANBased Indoor Positioning Scheme," *Proc. Intl Symp. World of Wireless, Mobile and Multimedia Networks*, pp. 1-8, June 2008.

[11] C. Feng, W. Au, S. Valaee, and Z. Tan, Received signal strength based indoor positioning using compressive sensing, *IEEE Transactions on Mobile Computing*, no. 99, Oct. 2011.

[12] AW.Tsui, YH.Chuang, HH.Chu,"Unsupervised learning for solving RSS hardware variance problem in WiFi localization" *Mobile Networks and Applications, 2009 - Springer*