香港浸會大學
HONG KONG BAPTIST UNIVERSITY

計算機科學系
Department of Computer Science

Report production coordinator: Claudia Chui, claudia@comp.hkbu.edu.hk, +852 3411 7079
Report Web Site: http://www.comp.hkbu.edu.hk/tech-report

Location Estimation via Missing Value Support Vector Regression

Zhi-li Wu[1], Chun-hung Li[1], Joseph Ng[1] and Karl R.P.H. Leung[2]

COMP-03-020

Release Date: September16, 2003

[1]Department of Computer Science,
Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong
[2]Department of Information and Communications Technology,
Hong Kong Institute of Vocational Education

Abstract

Location estimation using the Global System for Mobile communication (GSM) is an emerging application where noisy power levels are measured from multiple channels to infer the location of the mobile receiver. While geometrical and communications models have been adopted to tackle this estimation problem, the terrain factors and power fluctuations have confined the accuracy of such estimation. Traditional learning algorithms cannot be readily adopted to this estimation problem as the power readings are associated with varying stations. In this paper, we model the location estimation problem as a machine learning driven support vector regression problem where the spatial coordinates of the location are regressed against the power measurements. We also propose ways to amend kernels due to severe missing of power measurements when using support vector regression methods. Our machine learning approach shows a superior performance to conventional algorithms.

Report Cover Designer: So Kin Ming

Location Estimation via Missing Value Support Vector Regression

Zhi-li Wu, Chun-hung Li, Joseph Ng and Karl R.P.H. Leung*
Department of Computer Science, Hong Kong Baptist University
*Department of Information and Communications Technology
Hong Kong Institute of Vocational Education

Abstract

Location estimation using the Global System for Mobile communication(GSM) is an emerging application where noisy power levels are measured from multiple channels to infer the location of the mobile receiver. While geometrical and communications models have been adopted to tackle this estimation problem, the terrain factors and power fluctuations have confined the accuracy of such estimation. Traditional learning algorithms cannot be readily adopted to this estimation problem as the power readings are associated with varying stations. In this paper, we model the location estimation problem as a machine learning driven support vector regression problem where the spatial coordinates of the location are regressed against the power measurements. We also propose ways to amend kernels due to severe missing of power measurements when using support vector regression methods. Our machine learning approach shows a superior performance to conventional algorithms.

Contents

I. Introduction to GSM Positioning

In 1996, Federal Communications Commission (FCC) directed the rules that wireless service companies should provide location identification for wireless emergency call, and the accuracy of location estimation should be within 125 meters [2]. Since the year of 2000, FCC has stipulated separate requirements for both terminal-based and network-based positioning methods. And the FCC rules revised that 67% of estimations should be less than 50 meters and 100 meters in terminal-based and network-based solutions, respectively [6]. While a terminal-based positioning solution based on modifications of handsets, SIM cards or both, thus, subscribers have to upgrade their mobile phones or renew SIM cards to enable this location service. By the contrast, network-based solution estimates location by upgrading the operator network instead of the handset, therefore, it can benefit all mobile phone subscribers. In July of 2002, this wireless Enhanced 911 rules has been revised again to improve the effectiveness and reliability of 911 service by reporting the telephone number of a wireless 911 caller and the location of the cellular phone within 50 to 100 meters in most cases [4]. Hence, there had been a lot of investigations on location estimations for satisfying the rules of Enhanced 911 in FCC.

In general, the terminal-based method can provide more accurate estimate than network-based method, which is also reflected in earlier FCC rules. However, it requires the modification of the handsets. From an engineering point view, it is not easy to enable the functionality of location estimate by modifying handsets, even the GPS technology, which is the most mature and desirable to be introduced to assist the mobile positioning, is also inaccurate for populated or in-door area. Moreover, it is not only an engineering concern to upgrade millions of handsets in operation already. Thus there is a need to develop location estimation algorithms for unmodified handsets.

There are several network-based technologies available for cellular phones positioning. These methods utilized the signal attenuation, angle of arrival (AOA), time of arrival (TOA), time difference of arrival (TDOA), enhanced observed time difference (E-OTD), and time advance (TA) to locate unmodified cellular phones. Generally, all of these methods are based on knowing some location-related information of reference points and then relating them to the location of the base station. For example, the time for the signals returned from a mobile to base stations is known in the TDMA system, this information can be the measurement for distances from the mobile to base stations, which can be further exploited for location estimation.

But for GSM network, to implement time-involved estimate algorithms such as TOA, TDOA and E-OTD, expensive clocks for precise time synchronization should be installed. The Cellocate is such a TDOA-driven system by using GPS time synchronization. Their field tests based on four base stations resulted in an accuracy of between 187 and 287m [7]. Even though GPS receiver can be used for synchronization purpose, it would put a heavy burden on the network operator.

Signal strength-based method is more economic and compatible for GSM network [8], [10], [5]. It estimates a mobile location by utilizing the strengths of a set of signals propagated from base stations to the mobile. The strength information can be easily returned from the mobile to network side. Plus several easily retrievable parameters of the mobile and base stations, the network side conducts the calculation and reports the location estimate. Since signal exchange within mobile communication networks are common and cheap, the signal strength-based location estimation avoids the problems of installing expensive components.

The current paper considers a novel application of the use of the GSM system for location estimation that relies only on the signal strengths. Thus there is no need for implementing new special-purpose equipments in GSM network to fulfill the location estimation task. And different from some traditional strength-based approaches that are hurdled by the complexity of building a proper signal propagation model and inversely deriving the position from noisy and incomplete signals, we consider the location estimation as a machine learning problem. Under the setup of solving the location estimation as a machine learning problem, the support vector regression (SVR) method along with the well-design of kernels used for SVR are intensively studied. This approach avoids not only the complexity of determining a proper propagation model by traditional geometric or statistical approaches, but also the inverse/inference problem to derive location from models. The SVR method, complemented by well-designed kernels, also shows great tolerance to the randomness and incompleteness of GSM signals.

This paper is organized as follows. Section 2 describes the difficulties of position estimation in the case that GSM signals are very random and incomplete and the reasons for choosing SVR as an empirical location estimation method. In section 3, we describe how the support vector regression can be used in GSM-based location estimation. In section 4, we discuss how a well-selected kernel affects the location estimation and how the pathological kernels due to signal missing can be tackled by using a special SoE kernel or compensated by a novel fingerprinting method. Section 5 and 6 contains simulation and real testing results by comparing our approaches with other methods. Finally, in section 7, we present our conclusion and point out some further research directions.

II. Difficulties of Signal Strength-based Position Estimate

A. The Location Estimation Task

Given a set of N GSM base stations (BS) which are located at $\{(\hat{x}_i, \hat{y}_i)\}_{i=1}^{N}$, the location estimation problem is to report the mobile station(MS) location (x, y) given the strengths of a set of signals transmitted from base stations to the MS.

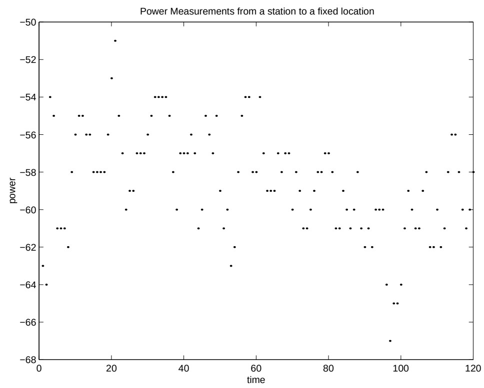

Fig. 1. Power Measurements from a station to a fixed location

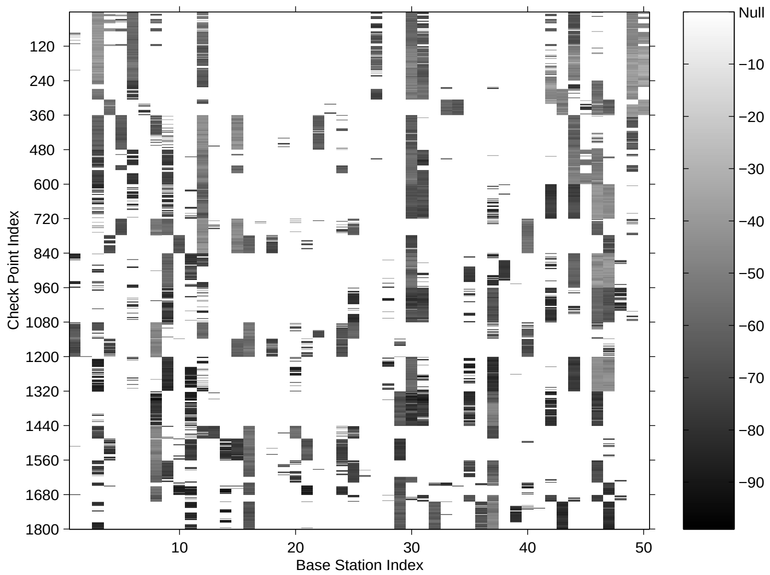

Fig. 2. Power Measurements from 45 stations to 15 fixed locations. 120 sets of measurements are taken for each location

B. Randomness and Incompleteness of Signal Data

As in previous discussions, it is favorable to use signal strength-based method for economy and compatibility reasons. But to achieve the accuracy required by the FCC rules, the random and incomplete characteristics of signal strengths are a great challenge for traditional location estimation algorithms.

To have an idea of the randomness of signal level from serving stations, a field test is carried out. The power measured from a fixed base station to a stationary point is taken over a 2 minute interval with a sampling period of one second. Figure 1 shows the power measured from the fixed terminal location which fluctuates randomly from a maximum of -51dbm to a minimum of -68dbm.

We now give a remark on the incomplete nature of signal strengths. Due to the GSM settings, only a small subset of signals propagated from base stations are large enough to be readable on a mobile handset.

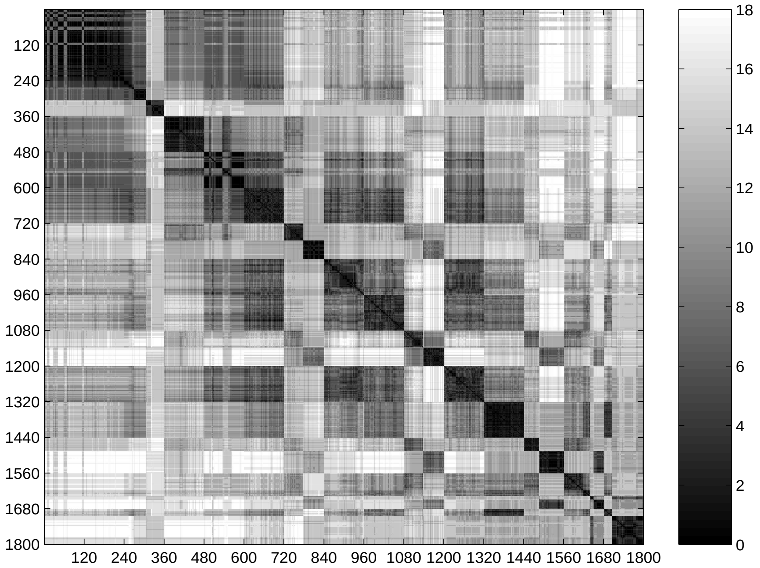

Fig. 3. Hamming Distance between all pairs of binary sequels which indicate the presence or miss of signals. The mean hamming distance is 11.

The subset, say n of N, usually ranges from three to nine [1]. The truncation threshold is not predictable although there may be a mechanical extreme for mobile handsets like -110 dBm. The only information we can know about the truncated signal readings is that they should at least be less than the smallest reading of the set of retrievable signals.

To have an idea of the incompleteness of signals, another field test is carried out. Fifteen check points are used for power measurement and 120 sets of signals are taken over a two minutes interval at per check point. By labelling all retrievable base stations in order, all data form a matrix with each column denoting the signal readings propagated from a same base station. Figure 2 shows the matrix in 2D image. Each blank entry with null reading indicates that the signal from the corresponding base station is too small to be retrievable at a fixed location. It is apparent that the matrix is very sparse when comparing the average retrievable stations at a single location to the total number stations once receivable at some locations, which is 9/50. This imposes difficulties upon some standard learning algorithms which are only effective to well-formed data vectors/matrices. Also it shows the signals fluctuate for each subset of 120 rows which in the diagram are shown between each pair of consecutive y-axis ticks.

The hamming distance can be used to measure the sparsity of the signal data matrix. The hamming distance measures the number of nonmatched bits for two binary codes. The Figure 3 shows the hamming distances between all pairs of binary sequels, while each binary sequel is converted from the signal strength vector with any effective entry being one and null entry being zero. The hamming distance matrix shows

[1]The 9 is the maximum number of signals retrievable by the Nokia handset (Model 6150), on which all our experiments and measurements are based

that the nonmatched entries for the current signal dataset can be up to 18, which are totally inconsistent concerning the average number of retrievable signals are 9. The mean hamming distance 11 implies that average matched entries are only (18-11)/2 = 3.5, which apparently demonstrates the sparsity of data vectors when concerning the whole vector length 50.

C. General Propagation Modelling V.S. Machine Learning Approach

The general signal propagation model approach derives the location from an parametric signal propagation model, which varies from an free-space model [5] to location-dependent HATA formula [10], even to an advanced one concerning random effects [11]. Since even the simplest signal propagation model is the nonlinear reaction between signals and location to be estimated, it is usually only possible to numerically approximate the position given the signals at a location. Such methods are claimed to only require off-site calibration when specifying the unknown model parameters. But it is always hard for these models to include enough environmental factors. To reach an desirable estimate accuracy such approaches should relax their model to be environmental-dependent, but environmental conditions are usually too volatile to be calibrated in an one-cut off-site manner, or too heterogeneous to be specified in a complete way.

The empirical location estimation methods require on-site calibration. Such approaches pay more attention to the empirical data once observed rather than an general model. Thus it may be the form of greatly simplified k-Nearest Neighbor rule [16] which assigns an unknown location to the weighted average of k-nearest locations observed before under a kind of definition for nearness and weight. This approach can be regarded as model-free, or at very most some of their concepts, like the nearness and weight in the k-nearest neighbor method, are partly involved with the assumption of propagation model. Or it is the case that some empirical location methods are model-supported in nature but the model are actually too sophisticated to be explicitly written down, like the neural network approach [16], which learns a network of highly generative models; or the support vector machine, which learns a simple linear model in feature space but produces a nonlinear output if projected back to the input space. The empirical location methods can achieve better accuracy in general at the cost of more real-time training/calibration data needed. But as in the previous discussion on the off-site calibration which does not absolutely demand less training data or possess longer period before re-specifying the model than the on-site calibration does in the location estimation case.

III. SVR Approach to GSM Location

A. Advantages of SVR as a Machine Learning Approach

Machine learning methods have been applied to indoor location estimation problem under a wireless local area network (LAN) setting [16], [17], [18]. However, in the indoor wireless LAN settings, signals are received from a small local area and less fluctuate due to less environmental change. Under this scenario, classical machine learning algorithms can be readily adopted for learning the locations given the readings from three or more different antenae. On the contrary, in the GSM location estimation problem signals may be transmitted from far away base stations and readings may be from various base stations. This scenario becomes a difficult missing value machine learning problem where traditional algorithms cannot be applied without modification.

The standard SVR approach in formulation is to build a linear regression model using the 1-norm loss function which is insensitive to small discrepancies. But more importantly the SVR approach is a kernel-based method which does not regress on the original input vector, but on its nonlinear expansion which is mapped from a kernel function to a very high dimensional feature space. While kernel-based learning methods have shown great success in data classification and clustering, regression estimation and function approximation. It is generally believed that data augmented to a higher dimensional feature space would be more separable, leading to Support Vector Classifier (SVC) and Support Vector Regression (SVR) and other kernel methods[9]. In next section, the mechanism of SVR will be further discussed.

Though powerful the SVR is, one issue arising from the severe incompleteness of GSM signal data must be addressed first. As explained in previous section, the the ratio of missing value can easily be over 80%. But conventional SVR approaches assume all values of the input vectors are known, which will not affect the evaluation of a kernel function and consequently will not significantly degrade regression performance even without any or with only a little effort put to handle the missing-value problem. Thus the GSM task has posed significant challenges to conventional kernel regression kernels. In this paper, we propose to use a Sum of Exponential (SoE) kernel which is robust to the severe missing-value phenomenon.

Before going to the detailed discussion of selecting a good kernel to tackle the missing value problem, we give the formal introduction to Support Vector Regression.

B. Introduction to Support Vector Regression

Following the common settings in machine learning, each set of N signal strength readings $\{s_q\}_{q=1}^N$ acquired at a MS at a specific time, are stacked into a columnwise input vector and denoted with a bold form $\mathbf{s}$.

For l sets of input vectors, we use $\mathbf{s}_i$ to represent the $i-$th vector and $\mathbf{s}_{iq}$ indexes to its $q-$th entry, namely, the $q-$th feature. And the target output is simply the coordinate (x_i, y_i).

Given a nonlinear mapping from input space $\mathbf{R}^N$ to feature space F such that $\phi : \mathbf{R}^N \to F, \mathbf{s} \mapsto \phi(\mathbf{s})$, together with a kernel function $k(\mathbf{s}_i, \mathbf{s}_j) = \phi(\mathbf{s}_i)^T \phi(\mathbf{s}_j)$, we then obtain a symmetric kernel matrix $\mathbf{K} = \{K_{ij}\}_{i,j=1}^l$, where $K_{ij} = k(\mathbf{s}_i, \mathbf{s}_j)$.

Consider the regression problem of approximating a linear function f that models each data $\phi(\mathbf{s}_i)$ to one of its corresponding coordinate, say x_i.

$$x_i = f(\phi(\mathbf{s}_i)) = \phi(\mathbf{s}_i)^T \mathbf{w} + b, \tag{1}$$

where the learning task now is to specify the constant value b, and the $\mathbf{w}$ which is an unknown vector with the same length as $\phi(\mathbf{s}_i)$.

By defining the linear $\epsilon-$insensitive loss function as

$$L(\phi(\mathbf{s}_i), x_i, f) = \max(0, |x_i - f(\phi(\mathbf{s}_i))| - \epsilon), \epsilon \geq 0, \tag{2}$$

where ϵ is usually a very small constant, the overall objective is to minimize the total loss

$$\min\{\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{l} L(\phi(\mathbf{s}_i), x_i, f)\}, (C \geq 0), \tag{3}$$

under the control of the parameter C, which is a user-specified nonnegative constant.

The primal optimization problem can therefore be defined as follows:

$$\min\{\tfrac{1}{2}||\mathbf{w}||^2 + C\textstyle\sum_{i=1}^{l}(\xi_i + \hat{\xi}_i)\}, \tag{4}$$

$$\text{s.t} \quad -(\epsilon + \hat{\xi}_i) \leq (\phi(\mathbf{s}_i)^T\mathbf{w} + b) \leq (\epsilon + \xi_i), \tag{5}$$

$$\epsilon \geq 0, \xi_i, \hat{\xi}_i \geq 0, i = 1, 2, \ldots, l. \tag{6}$$

The SVR maximizes the corresponding dual problem:

$$\max \quad \{\textstyle\sum_{i=1}^{l}(\alpha_i^* - \alpha_i)x_i - \epsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) \tag{7}$$

$$-\tfrac{1}{2}\textstyle\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(\mathbf{s}_i, \mathbf{s}_j)\}, \tag{8}$$

$$\text{s.t} \quad \textstyle\sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0, |\alpha_i| \leq C, i = 1, \ldots, l. \tag{9}$$

The $\mathbf{w}$ has the following expansion:

$$\mathbf{w} = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)\phi(\mathbf{s}_i). \tag{10}$$

Let $\beta_i = \alpha_i^* - \alpha_i$, the linear function f becomes

$$x_i = f(\phi(\mathbf{s}_i)) = \sum_{i=1}^{l}\beta_i k(\mathbf{s}_i, \mathbf{s}_j) + b. \tag{11}$$

To approximate y_i, the same methodology can be used. Also a sequence of more complicated regression outputs, such as the position denoted by the polar system, or the sequence of coordinates whose bases are related to each base station, can be designated for SVR. Some more efficient and robust output schemes are our further research topics.

An issue of SVR is to avoid overfitting the training data in case of simultaneously increasing the noise-tolerance, which is controlled by selecting parameters like ϵ, C and the one associated with the kernel function as shown in next section. In this paper, our experimental results are based on model selection through grid search upon these parameters.

IV. Missing Value Handling for Kernels and by Kernels

A. *Admissible Kernels*

In support vector classification and regression, a well-selected kernel determines how well the input data are mapped. Commonly used kernels are,

- Linear kernel

$$K(\mathbf{s}_i, \mathbf{s}_j) = \mathbf{s}_i^T\mathbf{s}_j,$$

- Polynomial kernel

$$K(\mathbf{s}_i, \mathbf{s}_j) = (\mathbf{s}_i^T \mathbf{s}_j + 1)^\gamma,$$

- Gaussian RBF kernel has following form:

$$K(\mathbf{s}_i, \mathbf{s}_j) = e^{-\frac{||\mathbf{s}_i - \mathbf{s}_j||^2}{2\sigma^2}} = \Pi_{q=1}^{N} e^{-\frac{(\mathbf{s}_{iq} - \mathbf{s}_{jq})^2}{2\sigma^2}},$$

where σ is a user-specified parameter. And $|| \cdot ||$ denotes the L_2 distance between two input vectors.

- First order ANOVA RBF kernel

$$K(\mathbf{s}_i, \mathbf{s}_j) = \sum_{q=1}^{N} e^{-\frac{(\mathbf{s}_{iq} - \mathbf{s}_{jq})^2}{2\sigma^2}},$$

where the ANOVA kernel can be derived under the framework of designing new kernels using ANOVA trick [21]. To our knowledge it has been implemented by [22] but without further references. Our SoE kernel shares the same base with the ANOVA RBF kernel but possesses further missing value handling mechanism.

B. Missing Value Handling

Conventional kernel functions evaluate each kernel value between two input vectors in a feature-to-feature manner. It thus always happens that the kernel value between two GSM input vectors can't be exactly evaluated due to missing feature values. To handle the missing value, one possibility is to ignore the missing value, simply by discarding the part of the feature-to-feature calculation involving the missing features. But this approach could easily make kernel value deviate from true value if too many feature-to-feature evaluation terms are thrown away. Thus the obtained kernel matrix might have to be compensated in some way. Or it might be necessary to select or design a more robust kernel function which is less sensitive to missing value. Yet another possibility is to impute missing values with some estimation, such as using the value zero or a fixed value with respect to their pre-confined bound, or an expectation for all or part of missing values. This approach generally performs well but has some uncertainty or difficulty. For example, replacing empty feature entries with a constant of zero is by default used in most real implementation [19] and just occasionally runs well. But in the GSM case the signal strength reading in dBm is always negative. Moreover, it might be futile to estimate an expectation for an area or a BS by assuming signal propagation is governed by a global model or mainly BS-dependent, without realizing it is actually highly specific to each location and affected by many local factors. The statistical approach for obtaining sufficient statistic of signal readings at each MS is extremely labor intensive and can only be conducted to a limited degree. Alternatively, a more sophisticated way is to put the problem under the general framework of completion of kernel [20] which utilizes some advanced techniques such as semi-definite program to deal with missing entries in a kernel matrix.

In this paper, the first two approaches are investigated. Namely, the SoE kernel which is robust and insensitive to missing values, an adapted RBF kernel which compensates missing value with fixed cost and another version of RBF kernel with zero filled in for missing feature entries.

C. Deal with Missing Value via SoE Kernel

A valid kernel feasible for support vector regression should normally meet the Mercer condition [1], which is critical to ensure the algorithm feasible in formulation and convergent to a unique optimal solution. Missing value usually makes a kernel matrix pathological. Thus it is essential to look for a missing value handling strategy to guarantee the validity of the repaired kernel matrix.

Filling-in strategy for original admissible kernel functions always results in legitimate kernel matrices. Thus a RBF kernel with zero filled in is immediately applicable without the concern of validity.

But it is not always valid for a kernel remedied by the strategy of excluding terms involving missing values. Now we show the adapted SoE kernel function produce valid kernel matrices.

We formulate the adapted SoE kernel function as,

$$k(\mathbf{s}_i, \mathbf{s}_j) = \sum_{q=1}^{N} \delta_{iq} e^{-\frac{||\mathbf{s}_{iq}-\mathbf{s}_{jq}||^2}{2\sigma^2}} \delta_{jq} \tag{12}$$

where δ_{ij} indicates whether the power reading transmitted from the $q-$th base station to the location (x_i, y_i) is truncated,

$$\delta_{iq} = \begin{cases} 0, \text{ if } \mathbf{s}_{iq} \text{ empty,} \\ 1, \text{ if } \mathbf{s}_{iq} \text{ otherwise.} \end{cases} \tag{13}$$

The kernel (12) can be derived from a series of combination of simple kernels shown as follows.

First by representing

$$(B_q)_{ij} = e^{-\frac{||\mathbf{s}_{iq}-\mathbf{s}_{jq}||^2}{2\sigma^2}}, q = 1, \ldots, l. \tag{14}$$

Each matrix $B_q \in \mathbf{R}^{l\times l}$ can be regarded as a Guassian RBF kernel formed with the pair of s_{iq}, s_{jq} as input vectors, which are a single entry of the original input vector $\mathbf{s}_i, \mathbf{s}_j$ respectively. And then by defining diagonal matrix $C_q \in \mathbf{R}^{l\times l}$ as

$$(C_q) = \text{diag}(\delta_{1q}, \ldots, \delta_{lq}), \tag{15}$$

the multiplication

$$D_q = C'_q B_q C_q. \tag{16}$$

is a kernel matrix ([1]-[Proposition 3.12.6]) with entries,

$$(D_q)_{ij} = \delta_{iq}(B_q)_{ij}\delta_{jq} \tag{17}$$

Finally, the summation of all kernel matrices D_q is also symmetric and positive semi-definite because

$$\eta'(\sum_{q=1}^{N} D_q)\eta = \eta'(D_1)\eta + \ldots + \eta'(D_N)\eta \geq 0, \eta \in \mathbf{R}^l. \tag{18}$$

So the proposed function (12) satisfies the requirement of a kernel.

D. SoE Kernel Compensated by Hamming Distance Metric

Hamming distance by its definition is the length of non-match bits, but it is also an Euclidian measure when the vectors are strictly constrained to be binary.

$$H(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) = \sum_{q=1}^{N} |\boldsymbol{\delta}_{iq} - \boldsymbol{\delta}_{jq}| = ||\boldsymbol{\delta}_i - \boldsymbol{\delta}_j||^2.$$

For such a valid metric measuring dissimilarity, we can convert it to a kernel function by exponentiating it.

$$K^{\text{HD}} = k(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) = e^{-\frac{H(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j)}{2\sigma^2}}.$$

Thus we can simply add the HD kernel to SoE kernel to compensate the discarded terms when evaluating SoE kernel values.

$$K = K^{\text{SoE}} + \lambda K^{\text{HD}},$$

where λ is a control parameter. The validity of the new kernel is straight-forward.

E. RBF Kernel Compensated by Fingerprinting Method

A fingerprinting method is proposed [ref...]:

$$K(\mathbf{s}_i, \mathbf{s}_j) = e^{-\frac{\lambda H(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) + \sum_{q=1}^{N} \delta_{iq}(\mathbf{s}_{iq} - \mathbf{s}_{jq})^2 \delta_{jq}}{2\sigma^2}},$$

where λ is a positive constant to compensate for the discarded terms due to the tensor indicators.

Since it is not easy to show whether it is a kernel, we empirically use it as a pseudo-kernel in SVR.

The fingerprinting method shares some characteristics with the traditional k-nearest neighbor method [16], but with the further improvement by introducing regularization term λ. The kernelled approach through SVR makes the solution for weighting more precise.

F. Compare SoE with RBF kernel

Many criteria, like VC-dimension and kernel alignment, can be used to evaluate which kernel is superior for a task. These statistics are usually mathematically profound but are relatively hard to evaluate. Practically we might be interested in directly comparing a pair of corresponding kernel values from two different kernel functions, shown by simple calculation or simple examples.

For RBF and SoE kernel, we have the following inequality:

$$\frac{1}{N} K(\mathbf{s}_i, \mathbf{s}_j)^{(\text{SoE})} \geq K(\mathbf{s}_i, \mathbf{s}_j)^{(\text{RBF})},$$

which can be proved from knowing the elementary term $\Delta_q = e^{-\frac{(\mathbf{s}_{iq} - \mathbf{s}_{jq})^2}{2\sigma^2}} \in (0, 1]$ and the Cauchy-Schwarz inequality,

$$\frac{1}{N} \sum_{q=1}^{N} e^{-\frac{(\mathbf{s}_{iq} - \mathbf{s}_{jq})^2}{2\sigma^2}} \geq \sqrt[N]{\Pi_{q=1}^{N} e^{-\frac{(\mathbf{s}_{iq} - \mathbf{s}_{jq})^2}{2\sigma^2}}} \geq \Pi_{q=1}^{N} e^{-\frac{(\mathbf{s}_{iq} - \mathbf{s}_{jq})^2}{2\sigma^2}}.$$

The above inequality shows that the kernel value evaluated from the SoE function is larger than that of using RBF. Moreover, considering the condition of equality holds, we know the difference between two kernel functions reaches the minima when all elementary terms Δ_q tends to be identical. Therefore in the

GSM case, the SoE and RBF kernel functions produce similar kernel values only when all pairs of their corresponding elementary terms are nearly the same.

Thus we know the SoE kernel function is more sensitive to each elementary difference term than the RBF kernel. In the GSM case, especially due to the side effect of missing value, we then have a first insight that the RBF kernel does not have the desired behavior compared with the SoE kernel.

Following concrete example further help us develop some understanding for the two kinds of different kernel functions.

we can focus on the case of two-dimension where we want to construct the similarity matrix between $\mathbf{s}_i = \{\mathbf{s}_{i1}, \mathbf{s}_{i2}\}$ and $\mathbf{s}_j = \{\mathbf{s}_{j1}, \mathbf{s}_{j2}\}$, , which are both complete input data.

Let

$$\Delta_1 = e^{-\frac{(\mathbf{s}_{i1}-\mathbf{s}_{j1})^2}{2\sigma^2}},$$

and

$$\Delta_2 = e^{-\frac{(\mathbf{s}_{i2}-\mathbf{s}_{j2})^2}{2\sigma^2}}$$

be the elementary terms of dimensional difference, the two kernel values are:

$$K^{\text{SoE}} = \frac{1}{2}(\Delta_1 + \Delta_2),$$

$$K^{\text{RBF}} = \Delta_1 \Delta_2.$$

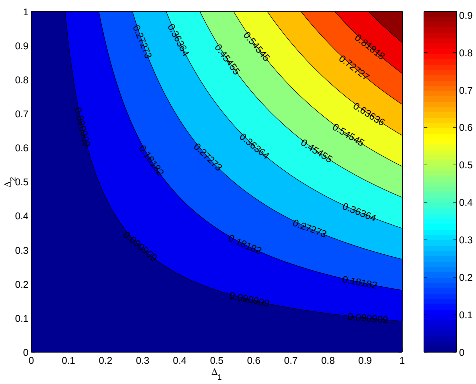

Fig. 4. RBF Contour.

From the contour graphs ($K(\Delta_1, \Delta_2) = z$) for RBF and SoE kernel function(Figure. 4, 5), we can observe that the differences of kernel value evaluation with respect to the two elementary difference terms Δ_1, Δ_2.

And it is easier to see their property difference from their substraction (Figure. 6).

From above graphs, we see that the closer the two elementary terms Δ_1, Δ_2 are, the less the kernel entries evaluated from the two kernel functions are different; But it is more important to notice that the SoE outputs a much larger kernel value if the two difference terms Δ_1, Δ_2 are more asymmetrical, which is consistent with the GSM case that a significant difference term Δ_i with respect to the $i-$th BS will dominate the decision of how near a pair of mobile stations are.

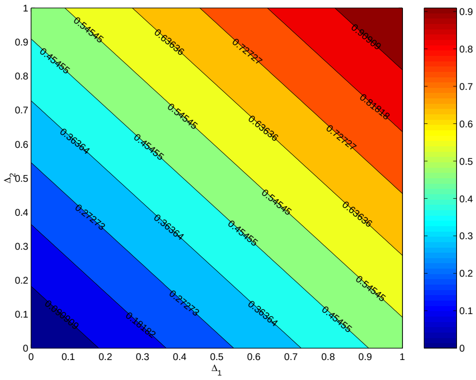

Fig. 5. SoE Contour

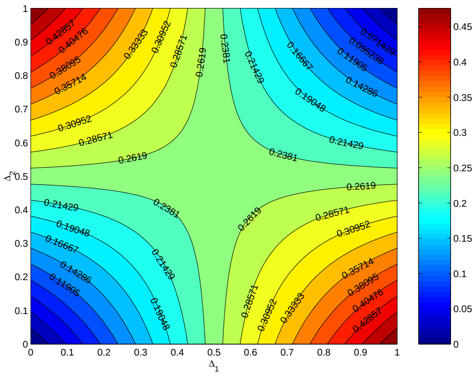

Fig. 6. Contour of $z = K^{\mathrm{SoE}} - K^{\mathrm{RBF}}$

V. Test on Synthetical Data

A. Benchmark Methods

Several geometrical and statistical modelling approaches are evaluated in current paper for comparison. The standard Cell-ID method, the Central of Gravity and the Trilateration method [5] and the triangulation approach [10] are common examples. Some of these methods are tested in our experiments. Also a statistical modelling approach [11], which uses EM algorithm for specifying a HATA [8] alike model and then derives the location from the model by grid search.

As the counterpart of signal propagation modelling approaches, many empirical location estimation methods are available for testing. Since the kernelized fingerprinting approach is a more precise version of the classical k-nearest neighbor estimation, we do not separately show results for k-nearest neighbor approach but we conjecture that the fingerprinting approach should improve the results achieved by k-nearest neighbor method. The neural network is more or less comparable to SVR approaches, but it does not possess some properties the SVR has, like global optimum, and kernel-based formulism. Furthermore, no good missing value handling strategies for neural network has been proposed, we also skip the comparison between SVR and neural network.

B. Synthetical Data Generator and Experiment Setup

The propagation model in the Statistical Modelling Approach [11] is used as a synthetic data generator for location estimation. But the original model has not taken into consideration of terrain factors and occlusion effect. By placing some randomly generated walls on the testing area and associating each wall with a random occlusion factor, signals received by a mobile station from base stations are attenuated by walls along the transmission paths, signals attenuated to be less than a threshold are simply truncated as null readings.

The simulated testing field has an area of 3000x3000 m^2 and with 20 base stations. The number of retrievable signal readings at each fixed position is 8. A number of walls, differing from 0 to 30, each with a random occlusion factor between 0 and 30 dbm , are randomly generated. Under each setting of walls, signals generated for 200 positions are used for training, 50 for testing. And 5 repeated runs are conducted for final report.

The EM algorithm in the statistical modelling approach converges after 400 iterations. After recovering the parameters for the propagation model, the testing area is divided into square meshes with size $50m \times 50m$, and then the weighted sum of a set of square centers with most largest likelihood are used as the expectation of the location of a testing point.

Parameters in support vector regression are specified in the following manner. The parameters like σs in SoE and RBF are specified by searching on a reasonable range and then are set to those achieving optimal testing results. The λ in RBF-fingerprinting are searched from the range of [0,1] in the same way. And other parameters are fixed at an empirically favorable value: c = 1000, epsilon = 0.01. Though it may be argued that the parameters are found in a over-learning manner. It is still convincing that the results under the optimal parameters settings can be regarded as best-case estimation.

C. Performance Analysis

The first three subplots in Figure 7 depict the test error of Cell-ID method, SoE, RBF and RBF-fingerprinting regression approaches as a function of the number of walls. And the fourth diagram shows the optimal parameters used by different SVR methods.

The Cell-ID method becomes more accurate when more walls are added. It might because the MS location is more determined by the locality of the serving station from which the strongest signals are received by a MS while relying less on other farther base stations whose signals vanish due to occlusion of more walls. The Cell-ID method runs fast for its simple estimation rule and does not require any training. But its performance is the worst among the testing algorithms.

The statistical modelling approach runs best when no wall is present. Because in such circumstance, the data is exactly generated by the parametric model to be learned, and using EM to recover several simple parameters is feasible, though computationally intensive. When the walls are gradually added, the

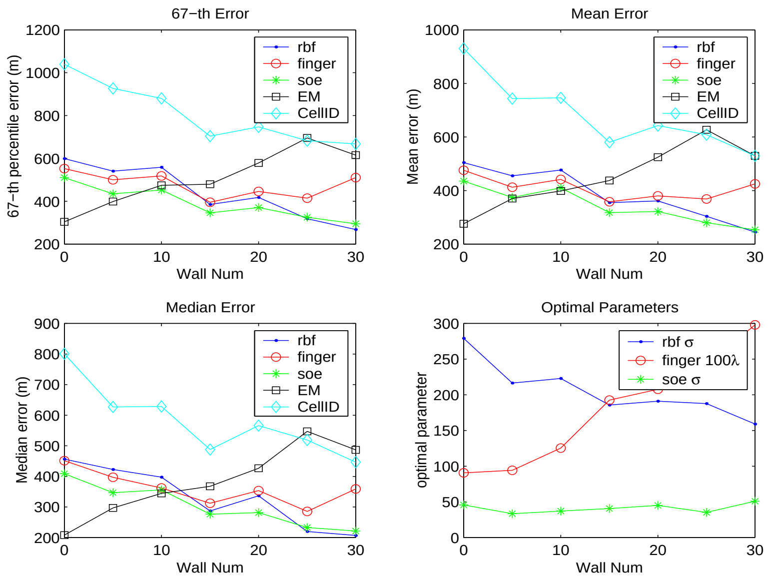

Fig. 7. Synthetical Test: Estimate Error as a function of the number of walls

statistical modelling approach deteriorates, and steadily degrades to be as worse as the Cell-ID method. Thus it can be concluded when the environment factors, such as walls and buildings, are hard to calibrated and significantly affect signal propagation, the propagation model recovered by EM approach is actually incompatible with the underlying generating model.

In general, all support vector regression approaches perform better and better when more walls are added. And the RBFFinger and RBF perform slightly worse, but when wall number reaches 30, the RBF becomes as good as the SoE. Moreover, the support vector regression approaches with SOE or RBF-fingerprinting kernels favors a set of stable optimal parameters, that is, searching for the optimal parameters of such SVR approaches will not cost too much. Thus the performance of different algorithms would still be similar if more rigorous parameter searching strategies like cross-validation is employed.

VI. Test on Real Data

A. System and Experiment Setup

A number of field tests have been conducted in cooperation with local mobile service providers.

One field test is conducted at TsingYi,which is a mountainous terrain. The base stations, markers for training and testing and area properties are are shown in a pseudo-map in Figure 8. Within the testing area a total number of 45 base stations are receivable by 47 checkpoints. At each checkpoint 10 to 300 sets of measures are taken.

Another field location is selected to be Mong Kok, which is a metropolitan area that is densely populated by tall buildings. The density of the buildings along with the narrowness of the street create a lot of occlusion and reflection that interfere with the measurement of the GSM power levels. Fifteen checkpoints

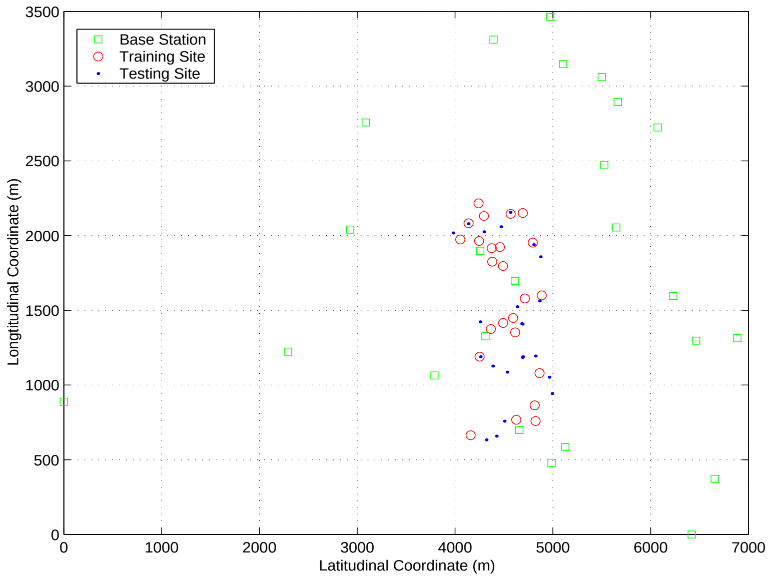

Fig. 8. Simulated Map for Tsing Yi and the relative placements of base stations and markers.

have been selected at street corners of the metropolitan area and 120 power measurements are taken from each checkpoint. When there is no positive lock-on from the GPS system the true locations of the checkpoints are obtained from GPS or by coordinates of the known landmark.

Fig. 9. The location of the 15 checkpoints in the populated area

The 15 checkpoints among the area are shown in Figure 9.

B. Training Multiple Sets of Readings at Each Fixed Training Position

Figure 10 shows the testing error as a function of the number of repeated data at each training position. Due to high variability of signal readings at a fixed position as shown in Figure 1, estimation error is inevitably large if only a single set of measurements is used for learning at each training position. Since it is very easy to obtain multiple sets of measurements within a very short duration at each handset, we test the performance when different number of repetitive training sets are used. As shown from the diagram,

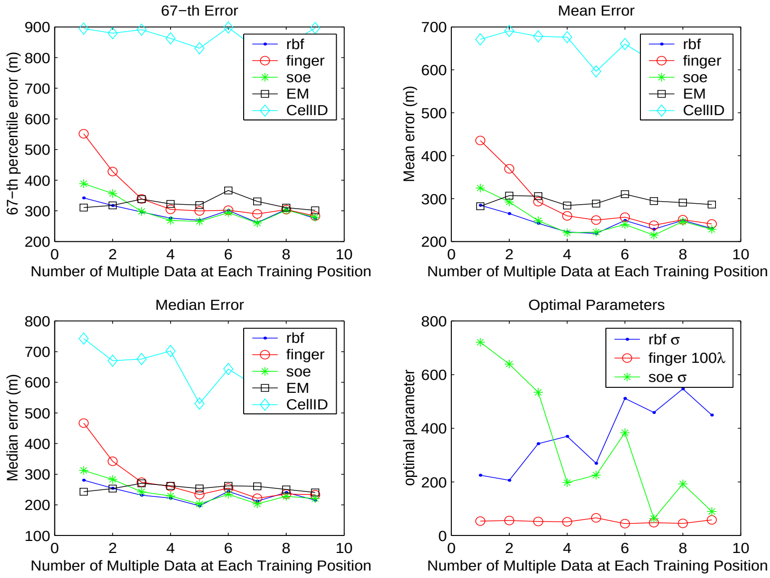

Fig. 10. Tsing Yi Field Test. Estimate Error as a function of the number of Multiple Data at Each Training Site, 40% markers are used as training.

adding 2 or 3 more sets of signal readings to each training position improve the estimate results for all algorithms except the Cell-ID method.

C. Change the Portion of Markers Used for Training

Figure 11 shows the testing error as a function of the portion of checkpoints used for training. The RBF SVR runs best by considering the testing error and the stability of optimal parameters. And the SoE also achieves better performance than the RBF-Fingerprinting and EM approach. The Cell-ID method performs worst.

D. Results for Mong Kok Area

The testing errors for Mong Kok are shown in Figure 12. When 9 of the 15 markers are exclusively used for training with 8 sets of signals for training per training marker, the estimate for each 120 sets of testing signals at the remaining 6 testing markers presents good consistency and accuracy. The training error is extremely low while one of the smallest mean testing error is 47 meters only. Further improvement can be obtained when multiple data are acquired within a short consecutive time using the average of each 120 sets of prediction as the final estimate, the estimate error further drops to 21 meters as shown in Table I.

Figure 13 shows the effect of changing the portion of markers for training when testing all kings of SVR methods. The SoE method outperforms other two approaches always.

Figure 14 shows the effect of increasing the number of multiple data for each training location when the portion of training markers is fixed at 6 (40%). Since the increment is very small by adding one more data per training site each time, it does not show the pattern of error drop. Though the curves averaged by 5

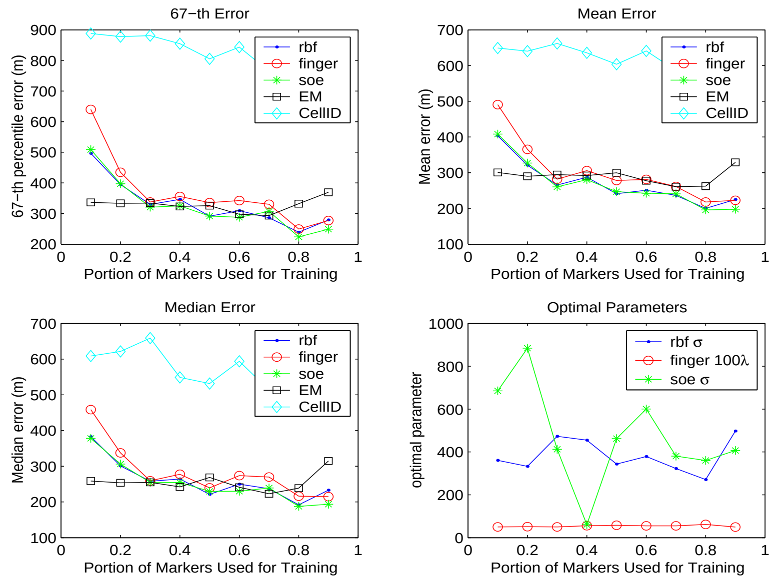

Fig. 11. Tsing Yi Field Test. Testing Error as a function of the change of portion of training examples. Each training site takes 4 set of readings, no requirement for testing data.

	CG (Weighted k-NN Cell-ID)	CT (Weighted 3-NN Cell-ID)	SoE SVR 1	SoE SVR 2
Mean Error(m)	85.29	95.18	47	21

TABLE I

AVE. ERROR IN POPULATED AREA BY CG, CT AND SoE SVR APPROACH. SoE SVR 1 IS THE SAME AS THE RESULT WITH (FIG.12). AND SoE SVR 2 FURTHER IMPROVES SVR1 BY PREDICTING EACH SET OF 120 MEASURES FROM A SAME CHECKPOINT TO ONE LOCATION

runs still zigzag, it is consistent with the results shown in above (Fig. 13) where the portion of training markers is set to 0.4.

The Cell-ID has not been tested for Mong Kok data, but the CG and CT methods [5], which can be regarded as the generalized weighted cell-ID version, is tested. When fixed at the same amount of data for training and calibration, the SVR approaches show superiority to the general modelling methods.

VII. Conclusion

In this paper, we have identified the GSM location estimation problems as a missing value data learning problem. The kernel based learning method and new kernel has been proposed for solving this estimation problem. The experimental results show that the support vector regression can predict the location of check points accurately using a very small set of training samples. Comparing with the statistical estimation method, the machine learning method is capable of providing better prediction involved in the environment in both synthetic simulation and field test in different locations in Hong Kong. One of the advantage of the machine learning method can be contributed to the capacity to adapt to local

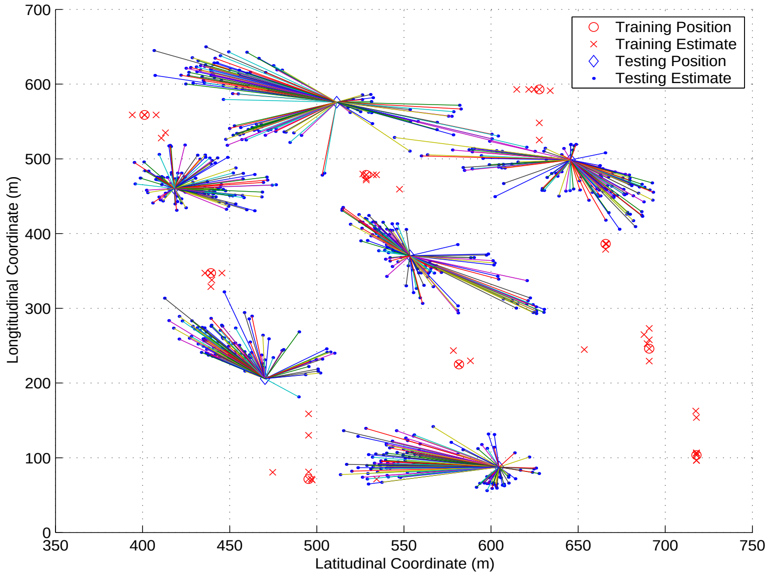

Fig. 12. Mong Kok Field Test. 8 of 120 sets of training readings from each of the 9 of 15 markers. All 120 sets of measures at the remaining 6 testing positions are used for test. Mean error = 47m

environment which is difficult to be modeled.

References

[1] Nello Cristianini, and John Shawe-Taylor, An introduction to Support Vector Machines and other kernel-based learning methods, pages 42-43, 1998.

[2] Federal Communications Commission.: "Revision to the commission's rules to ensure compatibility with enhanced 911 emergency calling system" July 1996, CC Docket No.94-102.

[3] Geoffery G. Messier, Michel Fattouche and Brent R. Petersen, "Locating an IS-95 mobile using its signal", *The Tenth International Conference on Wireless Communications (Wireless 98)*, vol. II, Calgary, AB, Canada, pp. 562-574, sponsored by TRLabs, the Communications Research Centre and IEEE Canada, July 6-8, 1998.

[4] `http://www.fcc.gov/911/enhanced`. FCC Enhanced 911 Services Review.

[5] Kenny K.H. Kan, Stephen K.C. Chan, and Joseph K. Ng, "A Dual-Channel Location Estimation System for providing Location Services based on the GPS and GSM Networks", to appear in *Proceedings of The International Conference on Advanced Information Networking and Applications (AINA),* March 27-29, 2003, Xidian University, Xi'an, China.

[6] `http://www.fcc.gov/911/enhanced/releases/factsheet_requi rements_012001.pdf`. E911 Phase II Decision.: "Fact Sheet of FCC Wireless 911 Requirements".

[7] Richard Klukas, Gerard Lachapelle, and Michel Fattouche, "Cellular Telephone Positioning Using GPS Time Synchronization", available at `URL:www.gpsworld.com/0898/0898innovations.html`

[8] Y. Okumura et.al. Field strength and its variability in vhf and uhf land mobile service, *Rev. Elec. Comm. Lab.*, pages 825-873, 1968.

[9] A. Smola, Regression Estimation with Support Vector Learning Machines, Technische University Mnchen, 1996.

[10] Svein Yngvar Willassen, "A Method of implementing Mobile Station Location in GSM", available at `URL:http://www.willassen.no/msl/bakgrunn.html`

[11] T.Roos and Petri and H.Tirri, "A statistical Modeling Approach to Location Estimation", *IEEE Mobile Computing,* 2002.

[12] Jeffrey Hightower and Gaetano Borriello, "Location Sensing Techniques", UW CSE Technical Report, 2001.

[13] Jim McGeough, "Wireless Positioning And LBS", Geomode Location Positioning White Paper, October 2001.

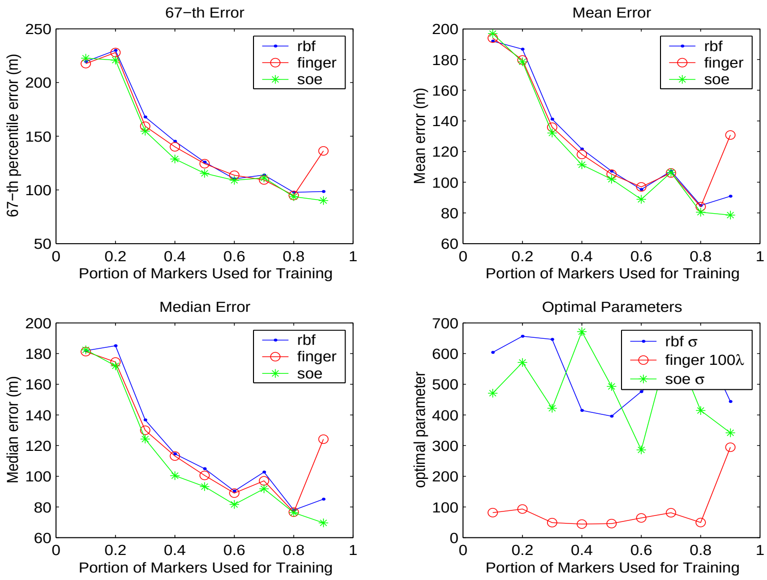

Fig. 13. Mong Kok Field Test Using SVR Methods. Testing Error as a function of the portion of markers for training. Each training site takes 8 set of readings, no requirement for testing data.

[14] Dragos Niculescu and Badri Nath, "Ad Hoc Positioning System (APS) Using AOA", Infocom03.

[15] B.W. Parkinson and J.J. Spilker, "Global Positioning System: Theory and Application", American Institute of Astronautics and Aeronautics, 1996.

[16] Roberto battiti and Alessandro Villani and Thang Le Nhat, "Neural network models for intelligent networks: deriving the location from signal patterns".

[17] M. Brunato and Cs. Kiss Kall, "Transparent location fingerprinting for wireless services", Med-Hoc-Net 2002, 2002.

[18] , "Roberto Battiti and et al.", "Statistical Learning Theory for Location Fingerprinting in Wireless LANs",2002.

[19] Chang, C.C. and Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

[20] T. Graepel: Kernel Matrix Completion by Semi-Definite Programming, ICANN 2002, 2002.

[21] Vladimir N. Vapnik: Statistical Learning Theory,John Wiley and Sons,1998.

[22] Stefan Rping, mySVM-Manual, University of Dortmund, Lehrstuhl Informatik 8, http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/, 2000.

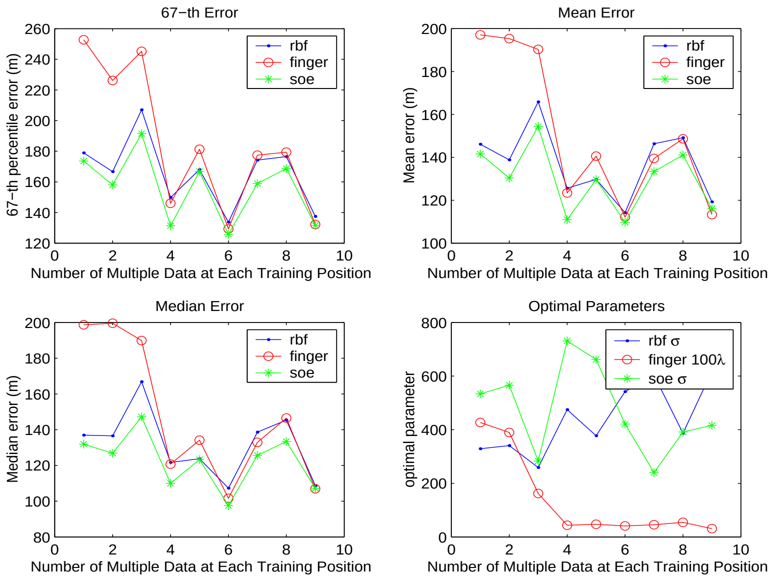

Fig. 14. Mong Kok Field Test Using SVR Methods. Testing Error as a function of the number of Multiple Data at Each Training Marker, 40% markers are used as training.