Sparse Gaussian Process for Spatial Function Estimation with Mobile Sensor Networks

Bowen Lu, Dongbing Gu, Huosheng Hu and Klaus McDonald-Marier

Abstract—Gaussian process (GP) is well researched and used in machine learning field. Comparing with artificial neural network (ANN) and support vector regression (SVR), it provides additional covariance information for regression results. By exploiting this feature, an uncertainty based locational optimisation strategy combining with an entropy based data selection method for mobile sensor networks is presented in this paper. Centroidal Voronoi tessellation (CVT) is used as a locational optimisation framework and Informative Vector Machine (IVM) is applied for data selection. Simulations with different locational optimisation criteria are conducted and the results are given, which proved the effectiveness of presented strategy.

I. INTRODUCTION

Wireless sensor networks (WSNs) are employed in various research fields which require to obtain sample data from a large scale of environment, such as forestry, meteorology, oceanography, etc [1][2]. For pollution monitoring, WSNs have been used for air and ocean environments in [3] and [4]. A sensor network is required to model the map of a spatial function in the environment. For a mobile sensor network, an effective locational optimisation strategy plays an important role on modelling performance.

Recently, kernel based regression methods are widely researched in machine learning field, including support vector machine (SVM) [5], relevance vector machine (RVM) [6], Kriging [7], and Gaussian process (GP) [8], etc. Some researchers found that these methods could convert from one to another under certain configurations [9]. Comparing among these kernel methods, the GP / Kriging gives additional uncertainty information for modelled distribution, which provides a criterion for locational optimisation.

Combining with the regression methods, various strategies are applied for locational optimisation in mobile sensor networks. Flocking algorithm was used with the spatial-temporal GP in [10]. Centroidal Voronoi tessellation (CVT) is another famous strategy introduced by Jorge Cortes from the view of computational geometry [11][12]. It was used with artificial neural network in [13].

Informative vector machine (IVM) is an entropy based method for selecting data from large number of samples, and was first presented by Neil Lawrence et al. in [14], [15], [16] and [17]. Their research shows that the IVM has a similar performance in computing speed and modelling accuracy with SVM.

In this paper, a framework of mobile sensor network for estimating a latent spatial function is given. More specifically, the CVT is used for optimising the sensor deployment; the GP estimates a latent spatial function and provides the uncertainty information for the model. In the CVT locational optimisation process, a combination of the mean and the covariance information is applied. Its performance is compared with using each of them alone. Potentially, a full data set is an ideal scenario for the modelling. However, using a full data set makes the computation of the GP model intractable. Therefore, the IVM is introduced to select a sub set with least information lost. The following sections are organised as below: section II-A gives a brief introduction to modelling the GP with a mobile sensor network, section II-B introduces the CVT with optimisation criteria configuration, and section II-C illustrates the principle of the IVM with data selection. Simulation results and analysis are given in section III.

II. MODELLING SPATIAL FUNCTION WITH WIRELESS SENSOR NETWORKS

A. GP Approach to Estimating the Latent Function

To estimate a latent spatial function \( f(x) \) in a 2D convex environment \( \mathcal{Q} \), a mobile sensor network with \( N \) sensors is deployed. The 2D coordinates of sensor \( i \in N \) are denoted by \( x_i \in \mathbb{R}^2 \) and its observation is \( y_i \in \mathbb{R} \). By collecting data from the sensor network, a training data set \( \mathcal{D} = [X, y] \) is constructed, where \( X := [x_1, x_2, \ldots, x_N]^T \) and \( y := [y_1, y_2, \ldots, y_N]^T \). With introducing a Gaussian noise \( \varepsilon_i \) to each sensor \( i \), an observation model is given in eq. (1):

\[
y_i = f(x_i) + \varepsilon_i
\]

where \( \varepsilon_i \sim \mathcal{N}(0, \sigma_n^2) \). Corresponding to Bayesian framework with the latent function \( f := [f(x_1), f(x_2), \ldots, f(x_N)]^T \), the observation \( y \) is the likelihood and its distribution can be illustrated as eq. (2):

\[
p(y|f) \sim \mathcal{N}(y; f, \sigma_n^2 I)
\]

where \( I \) is an identity matrix. The prior knowledge in the GP is defined by a kernel function \( K(x_i, x_j) \) and here is modelled by:

\[
K(x_i, x_j) = \sigma_n^2 \exp \left\{ -\frac{\|x_i - x_j\|^2}{2l^2} \right\}
\]
\( \sigma_f \) and \( l \) are two hyper-parameters, which can be modified online for controlling the amplitude and length scale of \( K(x_i, x_j) \). The prior distribution of the latent function is given in eq. (4):

\[
p(f|X) \sim \mathcal{N}(f; 0, K) \tag{4}
\]

Let a test point be \( D_\pi = \{x_\pi, f_\pi\} \). A joint distribution between the observation \( y \) and the GP prediction \( f_\pi \) is obtained in eq. (5):

\[
\begin{pmatrix} y \\ f_\pi \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ K_\pi \end{pmatrix}, \begin{pmatrix} K + \sigma_n^2 I & K_\pi \\ K_\pi & K_\pi^T \end{pmatrix} \right) \tag{5}
\]

where \( x_\pi \) is an arbitrary test point in the environment \( Q \), and \( f_\pi = f(x_\pi) \) denotes the predicted latent function value at \( x_\pi \). In addition, \( K_\pi, K_\pi \) and \( K_{xx} \) are shorthand notations, which denote \( K(X, X), \{X, x_\pi\} \) and \( K(x_\pi, x_\pi) \) respectively. The conditional distribution of \( f_\pi \) is obtained by conditioning eq. (5):

\[
p(f_\pi | f, X, x_\pi) \sim \mathcal{N}(\mu_\pi, \sigma_\pi^2)
\]

\[
\begin{align*}
\mu_\pi &= K_\pi^T \left[ K + \sigma_n^2 I \right]^{-1} y \\
\sigma_\pi^2 &= K_{xx} - K_\pi^T \left[ K + \sigma_n^2 I \right]^{-1} K_\pi
\end{align*}
\tag{6}
\]

\( \mu_\pi \) and \( \sigma_\pi \) from eq. (6) are the estimated latent spatial function mean values and uncertainty at \( x_\pi \). \( K_\pi \) and \( K_{xx} \) are controlled by the hyper-parameters \( \sigma_f \) and \( l \). In order to get the optimal values for \( \sigma_f \) and \( l \), the maximum log marginal likelihood is applied as eq. (7):

\[
\max_{\sigma_f, l} \{ \log p(y|X) \} \tag{7}
\]

where

\[
\log p(y|X) = -\frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} y - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{N}{2} \log(2\pi)
\tag{8}
\]

B. A Heuristic potential function for the CVT Localisation Optimisation

Centroidal Voronoi tessellation (CVT) is a gradient based locational optimisation method. To the CVT method for the sensor network in this paper, an arbitrary point from the environment \( Q \) is denoted by \( \hat{x} \), hence the Voronoi cell of sensor \( i \) is defined as:

\[
V_i = \{ \hat{x} \in Q | \| \hat{x} - x_i \| \leq \| \hat{x} - x_j \|, \forall i \neq j \} \tag{9}
\]

Before applying the CVT to the sensor network, a potential function needs to be defined. The latent function model generated from the GP contains two parts, mean \( \mu^* \) and covariance \( \sigma^* \). Some potential functions are available for various situations: using mean component \( \mu^* \) provides the smooth motion control result and an accurate local model, however it may be trapped at local mean maxima; using covariance component \( \sigma^* \) makes the sensor network cover as much area as possible, but it has lower accuracy on local details if the length scale is not large enough. Therefore, its modelling performance may end with a larger variance.

In balancing between them, we construct our potential function with the dot product of two components as equation (10):

\[
f_p = \mu^*_n \cdot \sigma^*_n \tag{10}
\]

where \( \mu^*_n \) and \( \sigma^*_n \in [0, 1] \) are normalised results from \( \mu^* \) and \( \sigma^* \) respectively. With this combination form, the sensor network can escape from local mean maxima.

For each Voronoi cell \( V_i \), its corresponding mass centre \( C_{V_i} \) is its optimal location:

\[
\begin{align*}
M_{V_i} &= \frac{1}{n_i} \int_{V_i} f_p(\hat{x})d\hat{x} \\
L_{V_i} &= \frac{1}{n_i} \int_{V_i} \hat{x} f_p(\hat{x})d\hat{x} \\
C_{V_i} &= \frac{L_{V_i}}{M_{V_i}}
\end{align*}
\]

C. Data selection: IVM

A full data set is constructed as \( D_f = \{D_1, \ldots, D_h\} \), where \( h \) indicates the length of \( D_f \). Normally, the criterion of selecting \( h \) depends on the changing speed of the latent spatial function. In balancing between the modelling accuracy and computing speed, an entropy based data selection method, the IVM [18] is introduced to select active data points from \( D_f \).

According to the principle of the IVM [18], it is necessary to update the posterior of the GP in a sequential mechanism. To achieve this goal, the IVM creates two index sets, active set \( I \) and inactive set \( J \). \( I = \emptyset \) and \( J = \{1, \ldots, h \times N\} \) are defined as their initial. The index of data points is selected one after another from set \( J \) to \( I \). An update form for computing the GP posterior estimation \( \hat{p}_i \) with \( \mu_i \) and \( \sigma_i \) is given (more details see [18]) as eq. (11), (12) and (13):

\[
\hat{p}_i \sim \mathcal{N}(f; \mu_i, \sigma_i) \tag{11}
\]

\[
\begin{align*}
\mu_i &= \mu_{i-1} + g_{n_i} \sigma_{i-1} e_{n_i} \\
\sigma_i &= \sigma_{i-1} + (g_{n_i}^2 - 2G_{n_i}) \sigma_{i-1} e_{n_i}^T \sigma_{i-1}
\end{align*} \tag{12}
\]

where

\[
\begin{align*}
g_{n_i} &= \frac{y_{n_i} - \mu_{i-1,n_i}}{\sigma_{i-1,n_i}^2 + \sigma_{i-1,n_i}} \\
G_{n_i} &= \frac{1}{2} \left( g_{n_i}^2 - \frac{1}{\sigma_{i-1,n_i}^2 + \sigma_{i-1,n_i}} \right)
\end{align*} \tag{14}
\]

\( n_i \) denotes the data point indices in set \( J \), and \( e_{n_i} \) is a unit vector choosing the \( n_i \)th element. \( \mu_{i-1,n_i} \) and \( \sigma_{i-1,n_i} \) are the \( n_i \)th element of \( \mu_{i-1} \) and the \( n_i \)th diagonal element of
respectively. With eq. (12), (13), the IVM can filter and select data points to active set $I$ while the GP updating the posterior $\hat{p}_i$ in sequential. The criterion of the IVM selection is the change of information entropy, and is defined as eq. (16) after the $i$th data point is included.

$$H_i = H(\hat{p}_i) := -\int \hat{p}_i(f) \log \hat{p}_i(f) df$$  \hspace{1cm} (16)

$\hat{p}_i$ is a Gaussian distribution, hence

$$H_i = \frac{i}{2} \log(2\pi e) + \frac{1}{2} \log |\sigma_i|$$  \hspace{1cm} (17)

The entropy change after the $i$th data point is selected into active set $I$

$$\Delta H_{i,n_i} = H_i - H_{i-1}$$

$$= \frac{1}{2} \log(2\pi e) + \frac{1}{2} \log |\sigma_{i-1}^{-1}|$$  \hspace{1cm} (18)

It should be noticed that $\Delta H_{i,n_i}$ is negative when the entropy is reducing. Therefore, the data points with the smallest $\Delta H_{i,n_i}$ values are selected by the IVM.

III. SIMULATION RESULTS

To simulate a 2D environment $Q$, an $1 \times 1$ rectangle area is chosen. A sensor network with $N = 15$ is randomly deployed at a corner of $Q$ ($x_i \in [0,0.2]$). The sensor noise level $\sigma_n = 0.01$, the hyper-parameters $\sigma_f = 0.01$ and $l = 0.01$ are initialised. The full data set length $h = 10$ and 15 active data points are set for the IVM data selection. A latent spatial function $f$ and its mesh graph are given in eq. (19) and shown in figure 1. An $100 \times 100$ evenly distributed grid mesh is configured for the environment $Q$.

The estimated latent spatial function model $\hat{f}$ is illustrated by two $100 \times 100$ matrices, $\mu^*$ and $\sigma^*$. Each value from the two matrices indicates the estimated mean value and the uncertainty level at a particular grid point respectively.

$$f = 1.9\{1.35 + e^{x-y} \sin[13(x-0.6)^2] \sin(7y)\}$$  \hspace{1cm} (19)

To illustrate the effectiveness of presented method, the simulation results are organised in three comparison groups and statistic data are collected from 100 tests. For the first group, the accuracy of estimated model from the IVM and the random selection are compared in figure 2. It can be observed that the IVM selection method provides a faster converging speed, smaller static error and standard deviation. In the next two groups, the IVM is employed as a data selection method to keep a fair comparison.

![Fig. 1. Latent function in environment](image)

![Fig. 2. Comparison between random selection and the IVM selection](image)

![Fig. 3. Comparison between $f_p = \mu^* \cdot \sigma^*$ and $f_p = \mu^* \cdot \sigma_{\text{norm}}^*$](image)
Introducing a data selection method, the IVM, a full data set is implemented without adding significant extra computational burden to the GP. The effectiveness of the IVM in data selection is proved by comparing with a random data selection process.

In the next step, a dynamic potential function could be studied to provide the flexibility to the sensor network with time variant or dynamic latent spatial function. Then the sensor network is able to change its locational optimisation criteria according to the real time information.

Acknowledgement: This research has been financially supported by EPSRC Global Engagements grant EP/K004638/1.

REFERENCES