Information Fusion 46 (2019) 193–205

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Approximate computational approaches for Bayesian sensor placement in high dimensions

Xiao Lin[a], Asif Chowdhury[a], Xiaofan Wang[b], Gabriel Terejanu[*,a]

[a] Department of Computer Science and Engineering, University of South Carolina, United States
[b] Department of Computer Science and Engineering, Xi'an University of Technology, United States

ARTICLE INFO

Keywords:
Mutual information
Bayesian inference
Canonical correlation analysis
Bayesian optimization
Low dimensional projection

ABSTRACT

Since the cost of installing and maintaining sensors is usually high, sensor locations should always be strategically selected to extract most of the information. For inferring certain quantities of interest (QoIs) using sensor data, it is desirable to explore the dependency between observables and QoIs to identify optimal placement of sensors. Mutual information is a popular dependency measure, however, its estimation in high dimensions is challenging as it requires a large number of samples. This also comes at a significant computational cost when samples are obtained by simulating complex physics-based models. Similarly, identifying the optimal design/location requires a large number of mutual information evaluations to explore a continuous design space. To address these challenges, two novel approaches are proposed. First, instead of estimating mutual information in high-dimensions, we map the limited number of samples onto a lower dimensional space while capturing dependencies between the QoIs and observables. We then estimate a lower bound of the original mutual information in this low dimensional space, which becomes our new dependence measure between QoIs and observables. Second, we use Bayesian optimization to search for optimal sensor locations in a continuous design space while reducing the number of lower bound evaluations. Numerical results on both synthetic and real data are provided to compare the performance of the lower bound with the estimate of mutual information in high dimensions, and a puff-based dispersion model is used to evaluate the sensor placement of the Bayesian optimization for a chemical release problem. The results show that the proposed approaches are both effective and efficient in capturing dependencies and inferring the QoIs.

1. Introduction

Sensor placement plays an important role in a range of engineering problems, such as grid coverage [8], target tracking [5,27], and monitoring atmospheric releases [20]. These problems may range across various subjects, however, they all share the same problem of inferring quantities of interest (QoIs) using measurement data. One such example is monitoring chemical release accidents where it is desirable to infer the release parameters such as location, strength and time to aid emergency responders. The release parameters are QoIs that can be inferred from sensor measurements. These type of sensor placement problems are model-driven and they require physics-based models to describe the spatial and/or temporal processes of the phenomenon of interest. In this context, QoIs are unknown model parameters and/or state variables, and the QoI inference problem involves running computer simulations and solving inverse and forward problems.

There are two types of approaches to solve the QoI inference problem: optimization methods and Bayesian inference. Optimization methods provide a single point estimate of unknown parameters by minimizing the discrepancy between model predictions and observational data. On the other hand, Bayesian inference incorporates prior information and provides a full posterior distribution over QoIs from which estimates with quantified uncertainties can be extracted. Sampling approaches such as Monte Carlo, Markov chain Monte Carlo, particle filters and ensemble Kalman filters are used to propagate uncertainties in forward problems and obtain samples from the posterior distribution calculated using Bayes rule. For a thorough review of both optimization methods and Bayesian inference, one can refer to [12]. In this paper, we focus on Bayesian inference.

Data collected from different locations may provide different amount of information towards QoIs. It is also impractical to place sensors exhaustively due to the high cost of installation and maintenance. Thus, sensors should be placed judiciously so as to maximize the information content. Sensor locations are commonly decided by running simulations and maximizing certain criterion. There are various choices such as mean square error (MSE), mutual information,

* Corresponding author at: Computer Science and Engineering, University of South Carolina, Storey Innovation Center, 550 Assembly Street, Columbia, SC 29208, USA.
E-mail addresses: lin65@email.sc.edu (X. Lin), asifc@email.sc.edu (A. Chowdhury), wangxfok@xaut.edu.cn (X. Wang), terejanu@cec.sc.edu (G. Terejanu).

https://doi.org/10.1016/j.inffus.2018.06.006
Received 7 March 2017; Received in revised form 18 June 2018; Accepted 19 June 2018
Available online 20 June 2018
1566-2535/

entropy, and invariants of information matrix such as A-, D-, and E-optimal design. Among all, mutual information is the most commonly used, since it is a natural measure of dependence between two random variables and the other criteria can be derived from it.

Ertin et al. [9] discussed maximum mutual information between sensor data and target state to decide which sensor should be queried for tracking the target. It was shown that maximum mutual information was equivalent to minimizing expected posterior uncertainty of the target state. This holds in the special case when the conditional distribution of the observable given the state is independent of the state as in the fixed additive Gaussian noise case. Krause et al. [18] discussed sensor placements for prediction problems where mutual information between the observed locations and unobserved locations is maximized in applications of monitoring temperature and precipitations. Wu et al. [29] tackled a similar problem in soil moisture. Instead of applying maximum mutual information strategy globally, locations were first clustered according to soil moisture content, and then maximum mutual information was used in each cluster to select sensor locations.

Although mutual information is a desirable criterion from a theoretical perspective, from a practical point of view it is challenging to estimate it directly from Monte Carlo samples. Some commonly used estimators include histogram based estimator, kernel density estimator, and k-nearest neighbor estimator (kNN). In their survey, Walters-Williams and Li [26] show that parametric estimation usually outperforms non-parametric estimation when data is drawn from a known family of distributions. But this is not the case in most practical problems. Khan et al. [15] compares different estimators and shows that the kNN estimator of mutual information developed by Kraskov et al. [17] captures better the nonlinear dependence than other commonly used estimators. The kNN estimator of mutual information, proposed by Kraskov et al. [17], is based on kNN estimator of entropy [16]. It is shown that by using matching distances in the joint and marginal spaces, the biases in entropy estimators could be canceled to provide an overall better accuracy.

The statistical errors of kNN estimators are approximately proportional to the dimensionality and inverse proportional to the square root of the number of samples [17]. This makes the estimation of mutual information challenging in high dimensions with limited number of samples as is the case with complex physics-based models. Nonetheless, in many applications, mutual information is adopted as a metric only for comparison purposes. In sensor placement applications, mutual information at different locations is compared to determine the sensor that can provide the most information for inferring the QoIs, thus, there's no need to know the exact value of mutual information as long as the correct comparison outcome is provided. As a result we can use alternatives that have consistent performance in such comparisons.

The contribution of the paper is two fold. First, we propose a lower bound of mutual information that is used as an alternative for comparison purposes. This is achieved by mapping the limited number of samples onto a lower dimensional space while capturing dependencies between the QoIs and observables. We then estimate a lower bound of the original mutual information in this low dimensional space, which becomes our new dependence measure between QoIs and observables. Second, Bayesian optimization [13] is introduced to facilitate maximizing the criterion over a continuous design space while reducing the number of lower bound evaluations. This strategy is inspired by the application of Bayesian optimization in experimental design by Weaver et al. [28], where the posterior variance of the failure-time quantile is maximized in an accelerated life test problem. Numerical results on both synthetic and real data are provided to compare the performance of the lower bound with the estimate of mutual information in high dimensions, and a puff-based dispersion model is used to evaluate the sensor placement of the Bayesian optimization for a chemical release problem. The results show that the proposed approaches are both effective and efficient in capturing dependencies and inferring the QoIs.

The rest of the paper is organized as follows. In Section 2, uncertainty modeling and Bayesian inference is introduced. The proposed sensor placement strategy is detailed in Section 3. Two numerical experiments are presented in Section 4 to show the consistency of the lower bound on both synthetic and real data and to demonstrate the efficiency of the proposed sensor placement based on the new lower bound and Bayesian optimization on a simulated chemical release accident. Finally, conclusions are given in Section 5.

2. Bayesian inference

In this section, we introduce Bayesian inference for solving inverse problems. In the Bayesian framework, uncertainties in state variables and parameters are described using probability distributions. Measurement data is used to update the knowledge of these quantities. It is desirable for the posterior distribution to have a small uncertainty and at the same time to capture the true value. The connection between these quantities and observation data is embedded in the mathematical model which describes the phenomenon of interest. Since the proposed approach in this paper is universal, an abstract model will be introduced first.

2.1. Uncertainty modeling

To set notation, consider the following abstract model:

$$\mathcal{R}(u, \theta, x) = 0 \tag{1}$$

$$y = \mathcal{Y}(u, \theta, x) \tag{2}$$

$$q = Q(u, \theta, x)\ . \tag{3}$$

Here, $\mathcal{R}$ is some operator, u is the solution or the state variable and θ is a set of parameters, which usually have a physical interpretation. x denotes the scenario which defines the problem being considered. In sensor placement, x usually refers to sensor locations. $\mathcal{Y}$ is a map from the solution to the prediction quantity y that can be compared with sensor measurements D. In addition, $\mathcal{Q}$ defines QoI which is denoted by q. In our problem, q can be θ or u or other quantities inferred from θ and u. However, no matter what q is, model parameters and state variables need to be known first. Then other quantities can be obtained through Eq. (3). Usually, θ and u are unknown or partially unknown, and need to be inferred from data. Let τ denotes unknown parts of θ and u. In this paper, Bayesian inference is carried out to solve the problem.

Because the observation data is noisy due to sensor imprecision, the measurement noise ϵ follows a known pdf $p(\epsilon)$ that is defined by the specifications of the sensors. This results in the following relation between the observable d and model prediction y.

$$d = y + \epsilon. \tag{4}$$

Finally, the relation between the observable d and model parameters θ as well as state variable u is given by combining Eqs. (2) and (4). This measurement model, Eq. (5) defines the likelihood function and Bayes rule can be used to update the knowledge of τ.

$$d = \mathcal{Y}(u, \theta, x) + \epsilon\ . \tag{5}$$

Since Bayes rule is used as the inference engine, then a prior probability distribution needs to be defined for τ, that is, $\tau \sim p(\tau)$. Note that the additive errors introduced in the previous equations are not a requirement; multiplicative errors or embedded errors are possible as